An integrative approach to studying speciation in a recent adaptive radiation

Michael Roy Montagu Lawson

PhD

University of York

Biology

February 2023

ABSTRACT

Alcolapia cichlid fishes are a small and exceptionally young species flock native to extreme soda lakes Natron (3 species) and Magadi (1 species) in East Africa. Previous research discovered narrow genomic divergence, overlapping niches and widespread interspecific gene flow, indicating that *Alcolapia* are at the nascent stages of speciation. We employed an integrative approach to study different mechanisms underlying their speciation. To investigate premating isolation, we characterised the strength of assortative mate choice between species using laboratory mating trials. Mating was similarly partially assortative between sympatric and allopatric species, suggesting that reproductive barriers can emerge or be maintained without recent gene flow. Next, to dissect the extent of dietary specialization among sympatric species, we characterised the cryptic diversity in the *Alcolapia* diet and gut microbiota using a metabarcoding approach. *Alcolapia* were found to lack some core microbial taxa found in other cichlids, but were enriched in extremophile taxa, suggesting that the hostile conditions may influence gut microbe communities. Sympatric species substantially overlap in their diet and gut microbial communities, and the small but significant differences indicated that while species differentially target certain trophic resources, they predominantly feed on the same sources using alternative foraging modes. Finally, we used genomic datasets to highlight extensive historical and ongoing gene flow between the species. Demographic models supported the paraphyly of the Natron species, with the *Alcolapia* common ancestor occurring before the Magadi-Natron split, suggesting a complex pattern of population structuring and hybridization even in this small radiation of fishes. Taken together, our findings indicate that a combination of ecological, sexual, and historical processes are involved in *Alcolapia* speciation. These results support a growing literature finding that reproductive isolation between sympatric species is driven by multiple mechanisms acting in concert.

Fighting male *Oreochromis* (*Alcolapia*) *alcalica*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

I dedicate this thesis to the memory of Corin Castle, who shared my passion for nature and science.

ACKNOWLEDGEMENTS

Completing this thesis took a long time and a lot of work, and over the years it would not have been possible without the help and contributions of many people whom I owe a huge debt of gratitude.

Firstly, I would like to thank my supervisor Kanchon Dasmahapatra for his continued guidance and expertise. He has helped me improve in all areas as a scientist and always had the trust, encouragement and patience for me to develop my own ideas.

I would also like to thank my co-supervisors Julia Day and Roger Butlin, and close collaborator George Turner. Julia's expertise on the system was invaluable, and I am very grateful to her for providing me with new opportunities and avenues for research. I thank Roger Butlin for valuable discussions and for helping me to broaden my understanding of the wider speciation field beyond just cichlids. I thank George Turner for hosting me at Bangor and providing his expertise on all things cichlid.

Much of this thesis builds on the work carried out by Antonia Ford, who has continued to be a generous source of knowledge even since leaving academia. In particular, I thank her and Julia for entrusting me with the metabarcoding project and offering additional insights beyond that found in the literature.

I greatly enjoyed my two stints at Bangor, including the opportunity to work at the cichlid facility, getting to know everyone at the ECH and Brambell buildings, and exploring the beautiful North Wales countryside. I owe a big thanks to Mike Hayle in particular, who not only provided me with somewhere to stay but was a massive help with the experiments and trying to rectify the fish disease issue. I am especially grateful for his help with the mate choice experiments, where he went above and beyond during the COVID-19 lockdowns. My thanks also go to Chloe Robinson, who carried out the pilot behaviour experiments, and Charlie Gregory, who was a great help in the cichlid facility.

I would also like to thank Elva Robinson and Dan Jeffares, who were on my thesis advisory panel at York; I appreciate their encouragement and different perspectives. In particular, I thank Elva for her insights on the behavioural aspects of the work.

Thanks to the NERC Environmental Omics facility at Sheffield for providing funding and resources for Chapter 3. I am especially grateful for the laboratory support I received from Paul Parsons and Gavin Horsburgh, whose help went far beyond my expectations.

I also thank our collaborators Joe Hoffman and Elke Hippauf at Bielefeld for carrying out the microsatellite marker analysis. While I haven't met them, I am grateful to Asilatu Shechonge, Dorothy Wanja Nyingi and all the people at TAFIRI and the National Museums of Kenya who helped facilitate the original fieldwork that the samples came from. My thanks also go to Domino Joyce and Alan Smith, who kindly agreed to house back-up populations of the fish at Hull and gave me a couple of guided tours of the cichlid facilities when I visited.

I am lucky to have been on the ACCE DTP and appreciate the opportunities for development and networking this gave me, including the excellent annual conferences and social events. ACCE funded my three-month-long placement at Natural England, which I benefitted greatly from thanks to Silas Walton and the freshwater specialist team's insights and guidance. I am also grateful for the opportunity I was given to present at the Speciation in Ancient Lakes 9 conference in 2022.

I thank all my friends, labmates and officemates from J1 and LCAB at York for their friendship and help over the years. Josie taught me to do DNA extractions, Lucie and Bruna helped a lot with analysis and stats, and Lewis passed on his *Alcolapia* knowledge. I am also grateful to everyone else in the Dasmahapatra and Ferrari groups, including Melissa, Edd, Alex, India, Julia, Chris and Yacine for all their helpful feedback during lab meetings.

I am incredibly grateful to my parents, Charles and Sue, and my sister, Vicki, who have always supported my interests; I could not have undertaken the last eight years of study without their help. Most parents would not allow 12 fish tanks when growing up! Thanks to my old school and university friends for all their friendship and support and providing a great source of entertainment when I needed to take a break. Most importantly, I would like to thank my partner Amy Newman, who has been a reliable source of encouragement and distraction when it's been most needed.

Lastly, I have appreciated exploring such an interesting topic and I want to thank the fish for being so weird and wonderful. After the past few years, I feel some kinship with *Alcolapia* as I now understand what it's like to survive in stressful situations!

DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. Where the work that contributed to this thesis was undertaken by someone else, this has been indicated below:

Live fish and voucher specimens used for genetic analyses were collected by Julia Day (University College London), Antonia Ford, Lewis White, George Turner (Bangor University), and Asilatu Shechonge (Tanzanian Fisheries Research Institute).

Live fish populations were maintained by Mike Hayle (Bangor University) and George Turner.

In Chapter 2, I carried out all behavioural experiments and the collection and processing of fish embryos and DNA swabs for the sympatric species experiment. Mike Hayle collected and processed fish embryos, fin clips and DNA swabs for the allopatric species experiment. I carried out the DNA extraction, PCR amplification and analysis of all samples. Microsatellite marker design was carried out by Kanchon Dasmahapatra (University of York) and Joseph Hoffman (University of Bielefeld), with sequencing carried out at the University of Bielefeld by Elke Hippauf.

In Chapter 4, library preparation for the whole genome re-sequenced dataset was carried out by Antonia Ford.

The data I present in Chapter 2 has been submitted to the journal “Frontiers in Ecology and Evolution” and is currently in review.

Michael Lawson

CHAPTER 1: GENERAL INTRODUCTION

1.1 Speciation

1.1.1 Speciation and the evolution of reproductive barriers

Speciation, the process whereby populations diverge to become separate species, is fundamental to the generation of biodiversity (Coyne and Orr, 2004). Conservative estimates place the total number of species on the planet in the order of 5 to 10 million (Mora *et al.*, 2011; Sweetlove, 2011; Costello *et al.*, 2013). Over the history of life on Earth, more than 99% of all species have gone extinct (Newman, 1997); speciation has therefore occurred orders of magnitude more times than the number of extant species. Since Darwin and Wallace introduced the theory of evolution by natural selection (Darwin and Wallace, 1858), understanding how new species arise has been a complex and contentious topic for evolutionary biologists (Coyne and Orr, 2004). Darwin described the origin of species as the "mystery of mysteries", and there are still large areas of disagreement about the underlying theories to this day (Darwin, 1859; Foote, 2018; Kulmuni *et al.*, 2020).

In sexually reproducing organisms, the concept of reproductive isolation, whereby diverging populations are prevented from producing offspring, is critical to the study of speciation (Dobzhansky, 1937; Mayr, 1942). Reproductive isolation can arise through a variety of mechanisms which are often visualised as reproductive barriers or isolating barriers to gene flow (Coyne and Orr, 2004). These can be classified into two broad categories, prezygotic and postzygotic barriers, which occur before and after fertilisation respectively. Prezygotic barriers include factors preventing mating from occurring, such as spatial and temporal isolation, differences in behaviour, incompatibilities of reproductive structures, and immigrant inviability (Turelli *et al.*, 2001; Coyne and Orr, 2004; Nosil, 2012). Postzygotic barriers cause the reduced fitness of offspring and are either extrinsic, where isolation is dependent on the environment (e.g. the inability of hybrids to occupy an ecological niche) or intrinsic to the host (e.g. hybrid sterility or impaired development) (Dobzhansky, 1937; Presgraves *et al.*, 2003).

Under the biological species concept, separate species form when reproductive barriers strengthen or accumulate to the point that populations cannot produce viable offspring. However, the proclivity for even highly divergent populations to occasionally hybridise has led to more relaxed definitions of species, allowing for some gene exchange (Mallet, 1995; Wu, 2001; Coyne and Orr, 2004). Characterising the reproductive barriers between diverging populations can provide insights into both the extent of reproductive isolation and some of the causative mechanisms underlying divergence (Coyne and Orr, 2004; Butlin *et al.*, 2012).

1.1.2 The role of natural and sexual selection in speciation

1.1.2.1 Natural selection

Natural selection is the process that occurs when individuals with certain traits are more likely to produce offspring with higher fitness than individuals with alternative traits. Other than in cases of reinforcement, natural selection does not necessarily favour genotypes that prevent gene flow unless they provide an adaptive benefit (Butlin, 1987; Turelli *et al.*, 2001; Via, 2009). The process where ecologically-driven divergent selection creates barriers to gene flow between populations is generally referred to as ecological speciation (Schluter, 2001; Rundle and Nosil, 2005). In contrast to other modes of speciation that may involve stochastic processes (e.g. genetic drift), ecologically based speciation occurs as a direct result of different ecological conditions (Schluter, 2001; Faria *et al.*, 2014). Ecological speciation can promote reproductive isolation indirectly as a by-product of ecological adaptation. Alternatively, reproductive isolation can be driven directly by ecological adaptation in the case of reinforcement, whereby hybrids have lower fitness than parental populations, and selection strengthens reproductive barriers that reduce hybridisation rates (Hendry, 2009; Nosil, 2012). Indirect ecological speciation can occur with or without gene flow, although direct reinforcement mechanisms are contingent on some population contact (Rundle and Schluter, 2004).

A prime example of ecological isolation is through habitat partitioning, where populations exhibit preferences for separate habitats, thereby reducing heterospecific contact (Dres and Mallet, 2002; Coyne and Orr, 2004; Rundle and Nosil, 2005). Migrants between populations may be poorly adapted and have reduced survival, further strengthening pre-zygotic isolation. The onset of sexual isolation as a result of

habitat separation is very common in nature, with a clear example found in the coastal snail *Littorina saxatalis*, where two ecotypes associated with different shore levels exhibit size-based assortative mating (Cruz *et al.*, 2004).

This ecologically-based selection can occur in response to multiple different environmental components such as habitat structure, dietary resources, competition, predators and parasites (Maan and Seehausen, 2011). To verify if ecological speciation drove divergence in a species pair requires that natural selection drove both the variation in phenotypes and reproductive isolation. While ecological divergence and reproductive isolation are widespread in nature, there has been some difficulty in linking these two processes (Hendry, 2009; Faria *et al.*, 2017).

1.1.2.2 Sexual selection

Sexual selection is a form of natural selection where one sex exhibits preferences for specific traits in the other sex and competes with members of the same sex for access to preferred mates. Sexual selection has long been linked as a common cause for reproductive isolation in sexually reproducing organisms, as the process can generate barriers to fertilisation (Darwin, 1859; Ritchie, 2007; Kraaijeveld *et al.*, 2011; Dobzhansky, 1937). In Fisher's (1930) runaway process of sexual selection, the coevolution of female preferences and male traits can promote evolution in a population independent of environmental factors. Theoretically, this can occur as long as there is genetic variation in a male phenotype that is subject to female choice (Lande, 1981). Indisputable evidence for this male trait-female preference co-evolution was initially elusive, but comparative studies have observed patterns consistent with this process (Panhuis *et al.*, 2001; Ritchie, 2007; Kraaijeveld *et al.*, 2011; Mendelson and Safran, 2021).

In the presence of gene flow, sexual selection may drive speciation through reinforcement, where reduced hybrid fitness promotes assortative mating (Butlin, 1987; Ritchie, 2007). Reinforcement by sexual selection may lead to reproductive character displacement, causing sexual traits to deviate in order to lessen the probability of sexual interactions between species (Brown and Wilson, 1956; Crozier, 1974; Pfennig and Pfennig, 2009).

Sexual conflict may stimulate speciation if female preferences outweigh male competition (Parker and Partridge, 1998; Ritchie, 2007; Gavrilets, 2014). For instance, if female fitness is reduced in response to a phenotype that benefits males, then male phenotypes will be subject to sexual selection, while female phenotypes are subject to natural selection. This can cause antagonistic sexual coevolution reminiscent of Fisher's runaway process, which can lead to morphological elaboration and reproductive isolation (Parker and Partridge, 1998; Rice, 1998; Mendelson and Safran, 2021).

Sexual selection may also promote speciation through mechanisms other than female choice, such as via male choice and male-male competition (Tinghitella *et al.*, 2018; Mendelson and Safran, 2021). For instance, in the common wall lizard (*Podarcis muralis*, males found in a two-population contact zone prefer females from their own population and outcompete the other population for preferred females (Heathcote *et al.*, 2016).

Contrastingly, sexual selection can potentially hinder speciation in some contexts (Parker and Partridge, 1998). For instance, sexual conflict arising from intense male competition may prevent assortative mating (Parker and Partridge, 1998; Tinghitella *et al.*, 2018). In addition, mating preferences may also be affected by perceptual bias, or perceptual adaptations that evolved in an alternative context, leading to more unrestricted mate preferences (Basolo, 1990; Ryan and Cummings, 2013).

1.1.2.3 The combination of natural and sexual selection

Both ecological and sexual selection necessarily interact as the evolution of mating systems is not independent of the environment. Both types of selection may also reinforce or resist each other throughout divergence (Safran *et al.*, 2013). Indeed, the recent consensus is that most cases of sympatric speciation (section 1.1.3) are driven by a combination of both ecological and sexual selection (Maan and Seehausen, 2011; Safran *et al.*, 2013; Servedio and Boughman, 2017; Mendelson and Safran, 2021).

A prime example of an interaction that promotes divergence is when adaptation to differing habitats affects signals involved in mate choice (Maan and Seehausen, 2010, 2011). The relative roles that sexual and natural selection play throughout speciation remains an active area of enquiry, with some evidence that sexual selection may be

more important during the early stages of speciation (Kraaijeveld *et al.*, 2011; Servedio and Boughman, 2017; Mendelson and Safran, 2021).

1.1.3 The role of geography and gene flow in speciation

1.1.3.1 The geographic context of speciation

Geography affects migration between diverging lineages and has important influences on the patterns of selection underlying divergence (Coyne and Orr, 1997; Servedio, 2016; Rosser *et al.*, 2019). Geographic barriers, where populations become physically (or allopatrically) separated, are perhaps the most widely known barriers to gene flow (Coyne and Orr, 2004). Given enough time, divergence becomes inevitable in this context, as local adaptation and other stochastic processes eventually lead to incompatibilities between lineages (Turelli *et al.*, 2001). However, many species arise and persist with at least some range overlap, where the exchange of migrants between populations is possible. Species contact can therefore range from complete separation at one end of the spectrum (allopatry) to partial overlap (parapatry) and complete panmixia (sympatry) (Coyne and Orr, 2004; Mallet *et al.*, 2009; Seehausen and Wagner, 2014).

One of the most contentious concepts in speciation research has been sympatric speciation, the emergence of two or more species without any physical barriers to gene flow (Bolnick and Fitzpatrick, 2007; Coyne, 2007). Traditionally, sympatric speciation has been defined as speciation that occurs in the absence of geographic barriers, but according to some population genetics definitions, a lack of spatial differences is not always sufficient for divergence to be considered sympatric (Coyne, 2007; Fitzpatrick *et al.*, 2008). Some of the disagreement over whether a case of speciation should be considered sympatric likely comes down to the use of different definitions of the term (Butlin *et al.*, 2012). By extension, the lack of consensus on what defines a species further complicates this problem (Mallet, 1995; Coyne and Orr, 2004).

Coyne and Orr (2004) suggested that to qualify as having originated sympatrically, species must fulfil these four criteria: 1) be distributed sympatrically; 2) have a monophyletic origin; 3) exhibit complete reproductive isolation; 4) past periods of allopatry are ruled out. Theoretical models have found this strict form of sympatric

speciation to be possible (Dieckmann and Doebeli, 1999; Bolnick and Fitzpatrick, 2007), although in natural systems, unequivocally satisfying these conditions has been difficult to impossible, particularly with regards to discounting historical allopatry (Bolnick and Fitzpatrick, 2007; Butlin *et al.*, 2008; Foote, 2018). Some empirical cases of divergent sympatric speciation have been widely accepted in examples such as the apple maggot fly (*Rhagoletis pomonella*) (Feder *et al.*, 1988), crater lake cichlids (*Amphilophus* spp.) (Barluenga *et al.*, 2006) and Lord Howe Island palms (*Howea* spp.) (Savolainen *et al.*, 2006). However, subsequent analyses have found that none of these systems fulfil the strictest definitions of sympatry (Feder *et al.*, 2003; Butlin *et al.*, 2008; Babik *et al.*, 2009; Kautt *et al.*, 2016). More verifiable forms of speciation in sympatry include hybrid speciation (see section 1.1.3.2) and autopolyploidy (Abbott *et al.*, 2013; Bolnick and Fitzpatrick, 2007).

The possibility of gene flow presents a problem for speciation as it homogenises populations by breaking down allelic combinations associated with ecological selection or assortative mate choice (Slatkin, 1987; Foote, 2018). There have been multiple theoretical solutions describing the genetic mechanisms underlying speciation with gene flow (Bolnick and Fitzpatrick, 2007). Feder *et al.* (2012) presented a four-phase model of speciation with gene flow which involves direct selection, divergence hitchhiking of tightly linked regions, genome-wide linkage effects, followed by post-speciation divergence. However, these phases are probably more indistinct than is provided by this metaphor, as differentiation across the genome varies continuously over the course of speciation (Faria *et al.*, 2017).

Determining the mode of speciation between species pairs can be difficult, as their current distribution may not reflect their historical distribution. Depending on their location and degree of reproductive isolation, this may cause species to have the appearance of allopatric or sympatric origin, while the inverse may be true (Coyne & Orr, 2004). They may also have had more complex histories involving periods of isolation followed by secondary contact (Richards *et al.*, 2019). Genetic analyses are often used to infer the likely route of origin, although these can also create misleading signals, such as false monophyly if historical gene flow and incomplete lineage sorting are not accounted for (Kutschera *et al.*, 2014; Cai *et al.*, 2021). Complementing traditional analyses by reconstructing the demographic history of species, including

estimating divergence times and ancestral gene flow, has been a powerful approach for disentangling these modes of speciation (Faria *et al.*, 2017; Foote, 2018; Bolnick *et al.*, 2023).

1.1.3.2 Adaptive introgression and hybrid speciation

Hybridisation also has the potential to accelerate speciation by providing a source of novel genetic variation (Abbott *et al.*, 2013). In areas of population overlap, the introgression of adaptive phenotypes may initiate divergence if they become associated with traits associated with reproductive isolation. This has likely played a key role in the rapid speciation and adaptive radiation of many taxa (Abbott *et al.*, 2013; Marques *et al.*, 2019).

Hybrid speciation is a form of sympatric speciation through adaptive introgression, where hybrids exhibit reproductive isolation from both parental species (Mallet, 2007). Allopolyploid speciation is one form of hybrid speciation and involves changes in chromosomes that create postzygotic barriers between hybrids and parents. While widespread in plants (Jiao *et al.*, 2011), this form of near-instantaneous hybrid speciation is less common in animals (Mable *et al.*, 2011). Homoploid hybrid speciation occurs without changes in ploidy, but initiates the reproductive isolation of hybrids and parental species through allelic recombination (Coyne and Orr, 2004; Mallet, 2007; Mavárez and Linares, 2008). Prezygotic barriers can emerge if hybrids exhibit reduced preferences for parental mating phenotypes or are able to exploit niches unavailable to the parental species (Abbott *et al.*, 2013). Postzygotic incompatibilities may also arise due to the reshuffling of gene combinations or as a consequence of chromosomal rearrangements (Mavárez and Linares, 2008)

1.1.4 An integrative approach to studying speciation

The forces driving speciation are diverse, shift over time, and can oppose and reinforce each other (Stankowski and Ravinet, 2021; Bolnick *et al.*, 2023). Focusing on a single mode of selection (e.g. sexual or ecological) may cause researchers to overlook other processes that may contribute to divergence (Maan and Seehausen, 2011). Likewise, barriers to gene flow naturally vary as speciation proceeds, meaning that characterising reproductive barriers only provides a snapshot of current reproductive isolation (Butlin *et al.*, 2012; Stankowski and Ravinet, 2021). To add

further complexity, individuals within populations are usually genetically and spatially heterogeneous, and the degree of gene flow between them can vary, requiring studies to sample across the full distribution (Coyne and Orr, 2004). Consequently, studying how species originate is difficult because speciation itself is multi-faceted and variable across space and time.

Early studies into speciation mostly centred around the morphological and ecological differences between species. After the modern synthesis, the focus shifted more towards the evolution of reproductive barriers, which incorporated several fields such as behavioural ecology, developmental biology and population genetics (Coyne and Orr, 2004; Kraaijeveld *et al.*, 2011; Faria *et al.*, 2017). Since the development of low-cost, high-throughput sequencing technologies, the field has increasingly used genomic datasets, which have provided valuable insights into the genetic basis of speciation (Seehausen *et al.*, 2014; Campbell *et al.*, 2018). Crucially, the field of population genomics has allowed the inference of past and current processes involved in divergence (Luikart *et al.*, 2003; Campbell *et al.*, 2018). This has been particularly important for understanding speciation with gene flow (Feder *et al.*, 2012; Foote, 2018). However, genome data is limited without an understanding of factors such as ecology and sexual selection (Faria *et al.*, 2017). Hence, a greater understanding of speciation can be developed using a more integrative approach that employs the tools from these separate fields and focuses on all stages along the speciation continuum.

1.2 Cichlid speciation and adaptive radiations

1.2.1 Adaptive radiations of the East African Great Lakes

Adaptive radiation is the process whereby a lineage diversifies to produce new species with a variety of different ecologies and morphologies (Schluter, 2000). Since its inception, adaptive radiations have been a focal area of study in the speciation field. Adaptive radiations have several features that make them well-suited to study speciation mechanisms. Firstly, adaptive radiations can comprise multiple species at different levels of reproductive and spatial isolation, meaning that divergence can be studied along the full extent of the speciation continuum and under different geographic contexts. They are also localised to specific regions and can occur rapidly (Givnish, 2015), making inferences on the causative processes underpinning

divergence more tractable compared to more distantly related lineages that are separated over larger spatial distances. Finally, they can contain repeating patterns of the same processes, which can act as natural replicates, providing a powerful approach for assessing the generalisability of different speciation mechanisms (Gavrilets and Losos, 2009; Salzburger *et al.*, 2014; Stroud and Losos, 2016).

Alongside examples such as Darwin's finches (*Geospiza* spp.) (Grant, 2017), three-spined sticklebacks (*Gasterosteus* spp.) (Schluter, 1993), *Anolis* lizards (Losos, 2011) and *Heliconius* butterflies (Merrill *et al.*, 2015), the adaptive radiations of cichlid fishes are model systems and have generated considerable research interest (Turner *et al.*, 2001; Kocher, 2004; Seehausen, 2006; Salzburger, 2018; Santos *et al.*, 2023).

Cichlids are a diverse and large family of fishes found predominantly in Africa and South and Central America. Consisting of around 3000 species, they are one of the largest vertebrate families (Seehausen, 2006; Salzburger, 2018). The most striking adaptive radiations are found in the East African rift lakes, where over 1200 species have evolved (Seehausen, 2006; Santos *et al.*, 2023). Lake Malawi contains a radiation of 500-860 species that diverged around 800 ka (Turner *et al.*, 2001; Brawand *et al.*, 2015; Malinsky *et al.*, 2018a). Meanwhile, Lake Victoria contains about 500 species, but the radiation is much younger, occurring in the last 15,000 years (Seehausen, 1996; Stager and Johnson, 2008). Lake Tanganyika is the oldest radiation (<10 Ma), contains about 240 species from 16 tribes and has the greatest morphological and genetic variation (Irisarri *et al.*, 2018; Ronco *et al.*, 2020; Svardal *et al.*, 2021).

In these systems, species have evolved from a small number of closely related ancestor cichlids and generated considerable variation in morphology, size, colouration, and life history (Brawand *et al.*, 2015; Santos *et al.*, 2023). A number of mechanisms have likely driven diversification in cichlids, including specific phenotypic and genetic features unique to cichlids combined with the availability of unexploited niches to colonising species (Wagner *et al.*, 2012; Brawand *et al.*, 2015; Seehausen, 2015; Wellborn and Langerhans, 2015).

1.2.2 Mechanisms of speciation in cichlids

1.2.2.1 Assortative mating

Cichlid fishes have provided some of the most convincing examples of sexual selection driving speciation (Ritchie, 2007; Mendelson and Safran, 2021). Related species often exhibit conspicuous variation in colouration that act as signals involved in species recognition and mate choice (Knight and Turner, 1999; Selz *et al.*, 2014; Rometsch *et al.*, 2020). Cues based on olfactory compounds and behavioural differences may play additional roles (Blais *et al.*, 2009; Van Steenberge *et al.*, 2022). Behavioural choice experiments have revealed that prezygotic isolation via female mate preferences is likely to be the most important isolating barrier between sympatric species (Seehausen and Van Alphen, 1998; Selz *et al.*, 2014; Rometsch *et al.*, 2020). However, mating preferences have also been suggested to play a lesser role compared to ecological factors at the early stages of speciation in the larger adaptive radiations (Muschick *et al.*, 2014).

1.2.2.2 Postzygotic barriers

In comparison to premating barriers, intrinsic postzygotic barriers are less common among closely related cichlid species (Rometsch *et al.,* 2020). Hybrid inviabilities generally take much longer to emerge between cichlids, with complete inviability even taking up to 4 million years to evolve (Stelkens *et al.,* 2010), which is more than the age of many cichlid adaptive radiations (Stager and Johnson, 2008; Brawand *et al.,* 2015). In a metanalysis of cichlid reproductive barrier studies, Rometsch *et al.* (2020) found that hybrid inviability was rare among closely related cichlid species, but sex ratio distortion was relatively common, with both increasing with increased genetic distance. Hence, hybrid inviabilities generally play a much lesser (but still a potentially important) role in early cichlid speciation than premating barriers such as assortative mate choice.

Extrinsic postzygotic barriers likely play a more important role than intrinsic barriers in early speciation (Coyne and Orr, 2004), but studies in cichlids have been relatively uncommon (Rometsch *et al.,* 2020). Accurately evaluating the relative fitness of hybrids requires comprehensive ecological studies, which can be challenging to collect, and likely explains the lack of research. Ecological postzygotic barriers have

been explored in *Astatotilapia burtoni*, where hybrids of lake and river ecotypes had reduced fitness in seminatural mesocosms (Rajkov *et al.,* 2018). In cichlids, an important extrinsic postzygotic barrier may be behavioural hybrid sterility, where hybrids have lower attractiveness to parental species (Coyne and Orr, 2004; Rometsch *et al.,* 2020). This has been experimentally documented in mate choice experiments using hybrids (van Der Sluijs *et al.,* 2008; Selz *et al.,* 2014) and may be an important mechanism of reinforcement in both sympatric speciation and during secondary contact. While hybridisation is frequently presumed to result in lower fitness, it can also promote speciation in cichlids by creating novel adaptive phenotypes (Keller *et al.,* 2013; Selz *et al.,* 2014).

1.2.2.3 Dietary and habitat partitioning

Ecological partitioning has been one of the primary driving forces for explosive diversification in cichlids (Kocher, 2004; Wagner *et al.*, 2012). Partitioning at different water depths and habitats has facilitated divergence through adaptation and spatial isolation (Genner and Turner, 2005; Albertson, 2008; Muschick *et al.*, 2012; Colombo *et al.*, 2016). Cichlids can also occupy a variety of trophic levels, and individual lakes can contain herbivorous, omnivorous, and carnivorous species (Ribbink and Hill, 1979; McKaye and Marsh, 1983; Bootsma *et al.*, 1996). High levels of specialisation are seen among Tanganyikan cichlids especially, where there are examples of planktivorous, insectivorous and scale eater species (Hori, 1993; Muschick *et al.*, 2012; Takeyama, 2021). Morphological diversification of trophic structures, including innovations to the pharyngeal jaw, have also helped enable the exploitation of different dietary niches (Liem, 1973; Hulsey, 2005).

Drivers of cichlid radiations also involve the interplay between ecological and sexual selection (Salzburger, 2009; Maan and Seehausen, 2011). In Lake Victoria, sympatric and closely related *Pundamilia* species display male colouration differences corresponding to water depth. Behavioural experiments and genetic studies identified that female preferences for male colour were based on the same opsin genes that conferred adaptation to different light environments found at different water depths (Seehausen *et al.*, 2008). These reproductive barriers are easily reversible among young species pairs when waters become turbid and diminish the transmission of sexual signals, further indicating that they emerge at the early stages of speciation

(Seehausen *et al.*, 1997; Maan *et al.*, 2010). Rapid speciation through sensory drive or via similar co-evolved ecological and sexual mechanisms may be a potent driving force in cichlid adaptive radiations (Maan *et al.*, 2006; Salzburger, 2009; Maan and Seehausen, 2011).

1.2.2.4 Diversification of the cichlid gut microbiome

Diet and habitat can lead to distinct gut microbial compositions, which can serve as a marker of ecological partitioning and adaptation to specific ecological niches (Ley *et al.*, 2008; Miyake *et al.*, 2015; Loo *et al.*, 2019). The gut microbiome may also influence speciation by providing a substrate for selection to operate (Shropshire and Bordenstein, 2016; Henry *et al.*, 2021). The study of the gut microbiome has extended into cichlid research, and sequencing of the gut microbiome of cichlids has found variation in the communities present between different lakes and at different trophic levels (Baldo *et al.*, 2019; Härer *et al.*, 2020). Echoing the findings in other taxa, these early studies have focused on the association of the gut microbiome with host diet, ecology and phylogeny (Ghanbari *et al.,* 2015). However, the role of the gut microbiome as a driver of speciation remains poorly understood (Rennison et al., 2019; Groussin et al., 2020; Miller et al., 2021).

1.2.2.5 Ancestral hybridisation and polymorphisms

For the last 20 years, the use of gene and genome sequencing by cichlid researchers has enabled investigation into the forces shaping cichlid speciation (Kocher, 2004; Brawand *et al.*, 2015; Salzburger, 2018; Svardal *et al.*, 2021). Using genetic datasets, hybridisation between current and ancestral lineages has generally been found to be an important contributor to adaptive divergence (Salzburger *et al.*, 2002; Genner and Turner, 2012; Weiss *et al.*, 2015). Genomic datasets have supported and elaborated on these findings by providing evidence that gene flow likely accelerated speciation in all three great lakes (Keller *et al.*, 2013; Meier *et al.*, 2017a; Malinsky *et al.*, 2018a; Irisarri *et al*., 2018; Marques *et al.*, 2019; Ronco *et al.*, 2020). Additionally, ancestral polymorphisms present during the early stages are increasingly recognised as important components of cichlid diversification (Brawand *et al.*, 2015; Meier *et al.*, 2017a; Irisarri *et al*., 2018; Marques *et al.*, 2019).

1.2.2.6 Geographic barriers and secondary contact

Spatial and geographic barriers within lake systems have also shaped the evolution of cichlid radiations. Habitat heterogeneity creates natural barriers between populations and patterns of intralacustrine parapatry and allopatry (Genner and Turner, 2005; Seehausen, 2015). For instance, many mbuna cichlids of Lake Malawi are restricted to shallow rocky habitats, which form a patchy distribution. Sub-populations between patches have been found to exhibit phenotypic divergence and local endemism (Rico and Turner, 2002; Genner and Turner, 2005).

Speciation has also been facilitated by fluctuations in lake levels which can create, destroy or merge habitat patches, causing species extinction, dispersal or isolation (Sturmbauer *et al.*, 2001; Ivory *et al.*, 2016). Likewise, allopatric divergence when species become temporarily isolated in peripheral habitats may initiate divergence (Barraclough and Nee, 2001) and has been observed in satellite lakes from the great lakes (e.g. Genner *et al.*, 2007; Tyers *et al.*, 2014). Reproductive barriers may emerge rapidly after isolation, with early signs of divergence recorded in a population that had only been separated for ~50 years (Moser *et al.*, 2018). The subsequent secondary contact between populations in peripheral habitat islands may accelerate speciation through reinforcement or adaptive introgression. The cycle of isolation and connectivity associated with these lake level fluctuations have been called “species pumps” for their role in promoting speciation (Sturmbauer *et al.*, 2001; Salzburger, 2009; Ivory *et al.*, 2016).

Riverine cichlids tend to exhibit lower diversity than their lacustrine counterparts, potentially due to greater temporal instability and lower ecological opportunity in comparison to most lake environments (Salzburger *et al.*, 2014; Seehausen, 2015). However, riverine cichlids have enabled initial lake colonisation and helped accelerate divergence in lake radiations during secondary contact events (Joyce *et al.*, 2011; Martin *et al.*, 2015; Poelstra *et al.*, 2018).

1.2.3 Small and recent cichlid radiations

Increasingly, small and young radiations from geographically isolated lakes have become popular testing grounds for studying the early stages of speciation. The simplicity of these radiations offers a more tractable study system for disentangling

various speciation processes (Barluenga *et al.*, 2006; Ford *et al.*, 2015; Poelstra *et al.*, 2018).

Notable examples of small lacustrine radiations come from crater lakes such as Apoyo and Xiloá in Nicaragua (Barluenga *et al.*, 2006), and Barombi Mbo and Ejagaham in Cameroon (Schliewen *et al.*, 2001; Schliewen and Klee, 2004; Martin *et al.*, 2015). Examples from freshwater lakes include Lake Mweru in Zambia (Stelkens and Seehausen, 2009) and Nabugabo in Tanzania (Bezault *et al.*, 2011), and the soda Lake Natron in Tanzania (Seegers *et al.*, 2001; Ford *et al.*, 2015). In addition, the rapid divergence (~1000 years) of species into distinct ecomorphs has been recorded elsewhere, such as in crater Lake Massoko, Tanzania (Malinsky *et al.*, 2015; Vernaz *et al.*, 2022).

Crater lake radiations are frequently cited as some of the most compelling examples of sympatric speciation (e.g. Bolnick and Fitzpatrick, 2007; Bord *et al.,* 2012; Foote, 2018). Due to the high rim of the caldera surrounding crater lakes, these bodies of water are effectively isolated from external rivers, which greatly reduces the likelihood of secondary contact events occurring. Furthermore, the tapering, conical shape of crater lake basins prevents the formation of separate bodies of water when water levels decrease, reducing the possibility of microallopatry occurring in the past (Barluenga *et al*., 2006; Foote, 2018). However, crater lakes still provide a variation of habitat niches, with some having significant depth gradients, which create distinct shallow-water and deepwater habitats for niche segregation to occur (Malinsky *et al.*, 2015).

The recent nature of some of these radiations from isolated lakes has helped shed light on the order of emergence of reproductive barriers. Ecomorphs from Lake Massoko showed evidence of divergence in genes associated with assortative mating preferences, suggesting an early role of sexual selection (Malinsky *et al.,* 2015). In Nicaraguan crater lakes and Lake Ejagham, ecological and sexual selection is suggested to have occurred simultaneously at the onset of divergence (Elmer *et al.,* 2009; Elmer *et al.,* 2010; Martin, 2013).

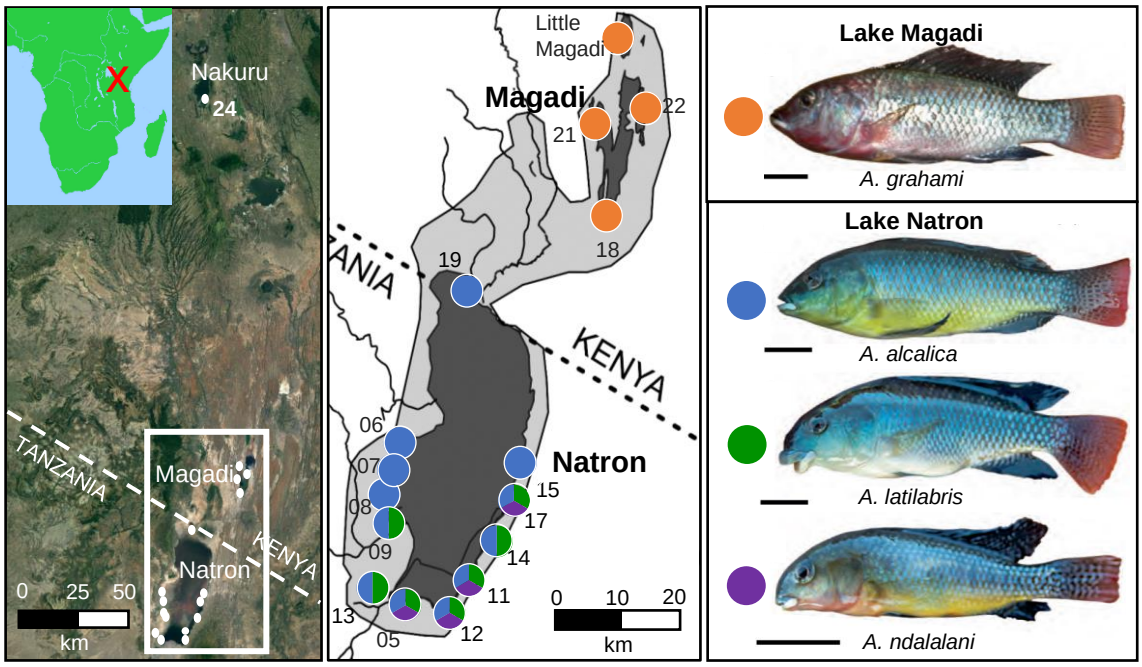


Figure 1.1. The *Alcolapia* radiation of soda lakes Natron and Magadi. Left panel: Map showing the distribution of *Alcolapia* in the East African rift valley with endemic lakes Natron (Tanzania) and Magadi (Kenya) and additional location (site 24) at Lake Nakuru (Kenya) contains *A. grahami* that were introduced from Magadi in the 1950's. Middle panel: schematic map of the *Alcolapia* radiation showing the distribution of species at different spring sites and lagoons (based on current sampling). Natron contains three described species (right panel: *A. alcalica*, *A. latilabris* and *A. ndalalani*); Lake Magadi and Little Magadi contain a single species (*A. grahami*). Coloured circles correspond to species present in the right panel (scale bar = 1cm). Dark grey areas denote the current extent of the lake, whereas the light grey area shows the approximate extent of Palaeolake Orolonga (from ~700 ka) (Williamson *et al.,* 1993). Photos, site information and schematic map are adapted from Ford *et al.* (2015).

1.3 The Alcolapia species flock

1.3.1 Introduction to the *Alcolapia* species flock

The *Alcolapia* species flock, also known as soda lake cichlids, are a small and recent adaptive radiation that have rapidly diversified in the extreme soda lakes Magadi and Natron in the East African rift valley (Trewavas, 1983; Ford *et al.*, 2019). *Alcolapia* are a subgenus nested within *Oreochromis,* a large genus with over 30 species that are endemic to lakes and rivers in Africa and the Middle East (Ford *et al.,* 2019; Eschmeyer *et al.,* 2022). *Alcolapia* are close relatives of *Oreochromis amphimelas*, which are

endemic to less extreme soda lakes in Tanzania (Ford *et al.,* 2019). They are also more distantly related to *Oreochromis niloticus*, commonly known as the Nile Tilapia, which is an important food fish that has been widely introduced to warm regions throughout the world (Ford *et al.,* 2019; Prabu *et al.,* 2019).

Lake Natron, which is situated in northern Tanzania, contains three described species: *Alcolapia alcalica, Alcolapia latilabris, Alcolapia ndalalani* (Hilgendorf, 1905; Seegers and Tichy, 1999), whereas Lake Magadi and its satellite lake Little Magadi, are situated approximately 25km over the border in Kenya, and contain a single species (*A. grahami*) (Boulenger, 1912; Seegers and Tichy, 1999; Figure 1.1). Both Natron and Magadi are volcanic lakes containing alkaline hot springs. The lake interiors are covered by thick trona deposits (sodium hydrogencarbonate) that are seasonally flooded (Kaufman *et al.*, 1990).

Alcolapia are restricted to the volcanic springs, streams and lagoons around the lake perimeter, where they can occur at high densities. In Natron, populations are most abundant in the south and southeast parts of the lake. While *A. latilabris* and *A. ndalalani* are restricted to these areas, *A. alcalica* populations have a wider distribution and can be found in sites to the far northern extent of the lake (Coe, 1966; Seegers and Tichy, 1999; Figure 1.1).

1.3.2 *Alcolapia* taxonomic placement and genetic relationships

Originally one species was described from Lake Natron (*Tilapia alcalica*) and one species from Lake Magadi (*Tilapia grahami*) (Hilgendorf, 1905; Boulenger, 1912; Woodhouse, 1912). Eventually, they were re-classified into the new genus *Oreochromis* and assigned the subgenus *Alcolapia*. Both were considered subspecies of *alcalicus*, *Oreochromis (Alcolapia) alcalicus alcalica* and *Oreochromis (Alcolapia) alcalicus grahami* (Trewavas, 1983). Seegers and Tichy (1999) carried out more extensive sampling across Natron and discovered two new species, *O. a. ndalalani* and *O. a. latilabris*. Based on significant divergence from other *Oreochromis* using mtDNA sequences, the *Alcolapia* sub-genus was also raised to genus rank (Seegers and Tichy, 1999; Seegers, 2008). A more recent analysis using both nuclear DNA and mtDNA markers found *Alcolapia* to nest within the *Oreochromis* genus, reverting

Alcolapia back to the status of subgenus (Ford *et al.*, 2019). For purposes of succinctness, we drop the *Oreochromis* name and refer to species as *Alcolapia*.

Early genetic studies identified population structuring between southern and northern Natron sites and between Magadi and Little Magadi (Seegers and Tichy, 1999; Wilson *et al.*, 2000, 2004), but using these small gene regions, the species relationships remained unresolved (Zaccara *et al.*, 2014). Later genomic studies using RADseq (restriction site-associated DNA sequencing) elucidated the population genetic relationships of *Alcolapia* (Ford *et al.*, 2015; Kavembe *et al.*, 2016). Echoing the findings of previous genetic studies, Ford *et al*. (2015) discovered extensive gene flow between sympatric Natron species. Likewise, for the Magadi populations, Kavembe *et al.* (2016) reported genetic variation localised by lagoon site, which corresponded with some fine-scale morphological variation. Demographic models also suggested recent divergence (<1700 generations) of these separate sub-populations, but with some migration post-split (Kavembe *et al.*, 2016).

Using the RADseq dataset generated by Ford *et al.* (2015), a comparative study investigating the extent of genomic variation within cichlid radiations found the Natron radiation had the lowest genetic diversity of all included studies (Svardal *et al.*, 2021). They found nucleotide diversity (π) and nucleotide divergence (d_{xy}) values significantly below other similar small and recent cichlid radiations, such as cichlids from Nicaraguan crater lakes and Barombi Mbo. Natron species had a nucleotide diversity (π) approaching 0.01%, among the lowest diversity recorded in any natural population and akin to the critically endangered Iberian Lynx (*Lynx pardinus*) (Abascal *et al.*, 2016; Svardal *et al.*, 2021).

While Ford *et al.* (2015) found exceptionally low interspecific genetic diversity, examination of variation across the genome also identified F_{ST} outliers, which are narrow regions containing many SNPs. These small regions of high diversity may contain genes involved in reproductive isolation and are speculated to be “genomic islands of divergence”, which are resistant to the effects of recombination and gene flow (Turner *et al.*, 2005; Ford *et al.*, 2015). However, it remains to be tested if these regions represent speciation islands or occurred due to alternate processes, as the genetic basis of these outliers has not been studied (Cruickshank and Hahn, 2014; Wolf and Ellegren, 2016).

1.3.3 Species descriptions and distribution

Seegers and Tichy (1999) provided descriptions of the defining features of each species along with morphological variants. *A. alcalica* and *A. grahami* share many morphological features, with conical heads and mouthparts typically terminal to subterminal, although there also exist populations of both species with upturned mouth morphology (Ford *et al.*, 2016; Kavembe *et al.*, 2016). *A. latilabris* have a derived subterminal mouth that is long and wide and with hypertrophied lips bearing dense and externally visible “teeth”. *A. ndalalani* has a subterminal mouth, but the head is short and narrow and with larger eyes (Tichy and Seegers, 1999; Figure 1.1).

Species also differ in male colouration: *A. alcalica* males have blue or bluish to yellow/golden at the breast, whereas *A. grahami* are generally grey or mauve to olive with a whitish to pink breast. *A. latilabris* are dark olive to light yellow on the body, with a lighter breast, throat and cheek. *A. ndalalani* males have intense orange colouration on the cheeks, throat and breast (Seegers and Tichy, 1999; Figure 1.1). In all species, females and juveniles exhibit more cryptic colouration and typically have a grey to olive colouration with faint vertical bars across the body (Seegers and Tichy, 1999).

All species are small relative to other *Oreochromis* species, rarely attaining a standard length over 6cm, but there is considerable variation in body size between sites (Seegers and Tichy, 1999; Ford, 2015; Maina *et al.*, 2019). Within-site comparisons found that *A. alcalica* and *A. latilabris* are consistently larger than *A. ndalalani* (Ford, 2015). However, all species attain significantly larger body sizes when kept in aquaria, suggesting that the lake conditions constrain their growth (Seegers and Tichy, 1999).

In addition to the four described species from Natron and Magadi, there is considerable polymorphism within species and between sub-populations (Tichy and Seegers, 1999; Ford *et al.*, 2015; Kavembe *et al.*, 2016). This includes variance in phenotypic traits such as mouth morphology (Ford *et al.*, 2016; Kavembe *et al.*, 2016) and also the presence of discrete morphs at certain sites (Tichy and Seegers, 1999; Ford *et al.*, 2015; Kavembe *et al.*, 2016). Notable examples include sympatrically occurring blue and yellow male *A. alcalica* morphs from site 11 and 12 (Ford *et al.*, 2015), upturned *A. alcalica* and *A. grahami* mouth morphs from site 15B in Natron and Little Magadi (Ford *et al.*, 2016; Kavembe *et al.*, 2016), discrete morph with

intermediate *A. alcalica* and *A. latilabris* phenotypes from site 14 (Seegers *et al.*, 2001; Ford, 2015), and melanistic morphs from hybrid site 17 (Ford *et al.*, 2015; Figure 1.1).

1.3.4 Adaptations of *Alcolapia*

Alcolapia survive in one of the most hostile environments inhabited by fish (Riesch *et al.*, 2015; Wood *et al.*, 2016). While other species can be found in similarly harsh habitats, it is the combination of extreme conditions that sets *Alcolapia* apart from others (Riesch *et al.*, 2015). Notably, this includes extremes of alkalinity (pH ~ 10.5), salinity (> 20ppt), and temperatures (30 – 42.8°C) (Ford *et al.*, 2015). The conditions are so hostile that when a closely related *Oreochromis niloticus* was placed in Lake Magadi it died in a matter of minutes (Wright *et al.*, 1990). Consequently, most research interest on *Alcolapia* has concentrated on their physiological adaptations for tolerating the extreme conditions present in the soda lakes (Randall, 1989; Narahara *et al.*, 1996; Bergman *et al.*, 2003; Kavembe *et al.*, 2015; Wood *et al.*, 2016; White *et al.*, 2022).

Observational studies have found *Alcolapia* to regularly graze in waters up to 40°C and temporarily pass through waters up to 44°C (Coe, 1966; Albrecht *et al.,* 1968). Prolonged temperatures at 42°C and abrupt temperature changes (from 22 to 36°C) are lethal (Trewavas, 1983). It has been found that *A. grahami* can survive in the hottest temperatures ever recorded for a fish, with one population having an upper critical temperature limit of 45.6°C (Wood *et al.,* 2016).

Between day and night, *Alcolapia* populations are exposed to dramatic fluctuations in oxygen levels (0.08–20 mg/L) (Johannsson *et al.*, 2014; Ford *et al.*, 2015). During the day, intense levels of algal photosynthesis create hyperoxic conditions before turning hypoxic at night as they respire. The arid conditions, elevation, proximity to the equator, lack of shade from terrestrial vegetation, shallow water depth (<1m), and high water clarity mean they are also subject to intensive ultraviolet radiation (Williamson *et al.*, 1993; Pörtner *et al.*, 2010; Johannsson *et al.*, 2014). The combination of elevated ultraviolet radiation, temperatures, pH and oxygen saturation generates reactive oxygen species (ROS), an irritant able to disrupt cell membranes (Johannsson *et al.*, 2014). *A. grahami* have been observed to aggregate in large groups, rise to the surface, and perform “air-breathing” behaviour during the mid-

afternoon when ROS levels are highest. It is speculated that this facultative air-breathing may be an adaptation to minimise ROS levels, which in Magadi reach some of the highest levels reported in the literature (Johannsson *et al.*, 2014).

Most fish are ammoniotelic and excrete nitrogenous waste in the form of ammonia (NH_4^+). This usually occurs via the passive diffusion of NH_4^+ to the water through a process of surface-layer acidification (Moreira-Silva *et al.*, 2010). The highly alkaline conditions of the lake make this diffusive excretion of ammonia impossible. As a result, *A. grahami* is the only known fish to be obligately ureotelic; instead of ammonia, *A. grahami* excretes urea (Randall, 1989; Wood *et al.*, 1994; Walsh *et al.*, 2001). Likewise, ureotelism is reported in *A. alcalica*, but in populations kept in aquaria over many generations, they also excreted some ammonia and had differential expression of the ammonia transporter genes compared to *A. grahami* (White *et al.*, 2020, 2022).

In *A. grahami* other innovations to survive alkaline conditions include elevated intracellular pH (Wood *et al.*, 1994) and separation of the stomach from the oesophageal channel, preventing alkaline water from interfering with gastric acids (Bergman *et al.*, 2003). One population of *A. grahami* have the highest metabolic rate of any fish recorded, approaching levels typically found in mammals (Wood *et al.*, 2016).

1.3.5 Sexual selection and mating systems in *Alcolapia*

Alcolapia, like other *Oreochromis*, are maternal mouthbrooders (Figure 1.2A). After fertilisation, the female cares for the eggs and fry until they are free swimming (Trewavas, 1983). There is little sexual dimorphism in size or morphology (Trewavas, 1983; Ford *et al.*, 2016), but they exhibit significant sexual dichromatism. Males display bright breeding colours when sexually mature (section 1.3.3; Figure 1.1), while females have more cryptic and subdued colouration with more prominent vertical bars along the body (Coe, 1969; Seegers and Tichy, 1999). Sexual maturation rates are rapid, with male breeding colours typically emerging around 30-40 days (Albrecht, 1968). Males also develop a prominent white lower lip, which appears to be a strong signal involved in male territoriality (Coe, 1966; Albrecht, 1968). Coe observed *A. grahami* males to attack models with white lips more frequently and with greater intensity than those without (Coe, 1966). Colouration intensity varies depending on the

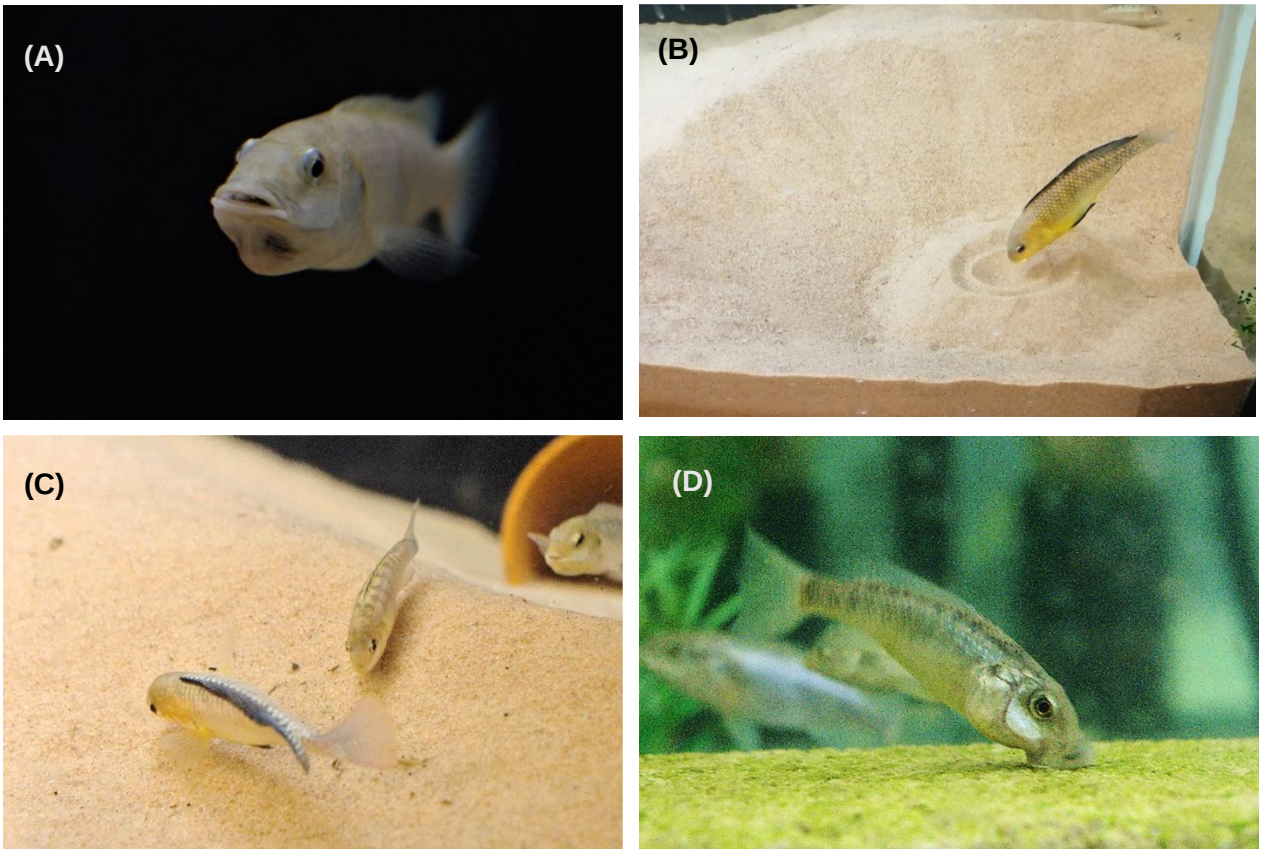


Figure 1.2. Reproduction and feeding behaviour in *Alcolapia* as documented in aquarium collections. (**A**) mouthbrooding of young fry by a female *A. alcalica*; (**B**) a male *A. alcalica* occupying a bower (breeding pit); (**C**) an *A. ndalalani* male courting a conspecific female at the periphery of a small bower; (**D**) Algae scraping by a female *A. latilabris* (Photographs: Michael Lawson).

position in the male dominance hierarchy, with subordinate *A. grahami* males observed to show subdued colouration in the presence of dominant males (Coe, 1969). Male *Alcolapia* also possess white genital papilla which are possibly “egg dummies” (Coe,1969; Trewavas,1983; *personal observation).*

As is typical of *Oreochromis*, mature males construct bowers, a type of courtship structure formed in the substrate (Trewavas, 1983; Figure 1.2B). These are generally found in the shallow water areas of the lagoon, where the substrate consists of sand or mud (Albrecht, 1968; Coe, 1969). In the vicinity of their bowers, males become highly territorial and perform both courtship and spawning behaviours (Coe, 1966; Figure 1.2C). Albrecht (1968) noted that *A. grahami* pits can have complex terrace-like structures with smaller, adjacent subsidiary pits. Where the substrate is rocky, pits sometimes comprise of a simple depression (Albrecht, 1968). Due to the high fish

densities, breeding pits occur close together, with Natron species pits separated by up to 60cm (Coe, 1969), whereas *A. grahami* pits can be much more concentrated (Coe, 1966), although the density of bowers varies according to site (Seegers and Tichy, 1999). In Natron, these leks can contain a mixture of males from all three Natron species (Seegers and Tichy, 1999).

Reproductive cycles are highly rapid in *Alcolapia,* with mature *A. grahami* females able to produce new broods approximately every two weeks (Coe, 1966). In Magadi, females have been found to deposit fry in small pools around the edges of the lagoon. These may act as nursery areas, protecting fry from higher temperatures and more saline-alkaline conditions, in addition to preventing predation by other cichlids (Coe, 1966; GD Kavembe and CM Wood, *personal communication* in Schagerl, 2016).

The promiscuous lek mating system and conspicuous male nuptial colours are suggestive of intense sexual selection operating in this system (Dale *et al.*, 2015; Henshaw *et al.*, 2016). Assortative mating has only been reported anecdotally in aquarium populations where some assortative mating was observed in a mixed species tank, although heterospecific courtship was also recorded in male *A. alcalica.* Hybrid *A. latilabris*/*A. ndalalani* were also generated, but only in the absence of conspecific males (Seegers *et al.*, 2001). While natural selection has driven the evolution of extreme adaptations in a short timescale, intensive sexual selection may similarly play an essential role in *Alcolapia* evolution and speciation.

1.3.6 Ecology and niche diversification of *Alcolapia*

The harsh conditions of the soda lakes prevent the colonisation of most organisms and *Alcolapia* are the only aquatic vertebrates present in lagoons and spring sites. Only a few invertebrate taxa are present, which mostly consist of dipterous larvae and copepods (Coe, 1966). The extreme conditions also inhibit the establishment of most vascular plant species, although some halophilic sedges and grasses are present around the littoral margins of some sites (Kipkemboi, 2016; Mwaniki *et al.*, 2019). Instead, the majority of primary production within springs and lagoons derives from algae, particularly cyanobacteria (Coe, 1966; Krienitz and Schagerl, 2016). On the surface, this depauperate environment appears to lack trophic complexity and a variety of ecological niches, yet sympatric Natron *Alcolapia* species possess unusually

divergent trophic morphologies not found in other *Oreochromis* species (Seegers and Tichy, 1999; Ford *et al.*, 2016).

Analysis of stomach contents has found all *Alcolapia* species to be predominantly algivorous (Coe, 1966; Ford *et al.*, 2016). Ford *et al.* (2016) investigated differences in the diet of Natron species. Observation of stomach contents, gut length measurements and stable isotope analysis uncovered considerable overlap between species, but with some subtle differences. Notably, *A. alcalica* showed evidence of feeding on more vascular plants than other species, whereas *A. ndalalani* and *A. latilabris* had substantial overlap (Ford *et al.*, 2016). Using the same approach, differences were also recorded between *A. grahami* from Magadi and Little Magadi, with populations of the latter also exhibiting upturned mouth morphology (Kavembe *et al.*, 2016). Upturned-mouth populations are speculated to feed on surface film or insects (Seegers and Tichy, 1999). However, these dietary analyses could not distinguish between algae taxa, of which there is considerable diversity in Natron and Magadi (Krienitz and Schagerl, 2016; Krienitz *et al.*, 2016), potentially missing cryptic dietary variation between species (Ford *et al.*, 2016).

Alcolapia are observed to graze almost continuously on cyanobacteria in the wild (Coe, 1966; Johannsson *et al.*, 2014), which is likely necessary for them to be able to maintain their high metabolic rates (Wood *et al.*, 2016). In Natron, each species is observed to graze differently depending on their trophic morphology (J.J. Day, *personal communication*, 21st November 2022). This can also be observed in aquarium populations. *A. latilabris* perform a distinctive algae-scraping feeding behaviour where their downturned mouths form a suction layer, allowing them to adhere to the substrate. Meanwhile, the “teeth” present on their hypertrophied lips are used to rasp at the algae present on the surface of rocks (Figure 1.2D). Alternatively, *A. ndalalani* browses on the substrate by picking off algae in more discrete motions (Ford *et al.,* 2016).

Cannibalism of offspring is common in *Alcolapia* (Coe, 1966; Philippart and Ruwet, 1982; Seegers and Tichy, 1999; Ford *et al.*, 2016), with males and non-breeding females observed to snatch fry and eggs from brooding females (Coe, 1966). As *Alcolapia* reproduce frequently, paedophagy could an important feeding strategy for some individuals (Coe, 1966; Trewavas, 1983).

Alcolapia appear to have fairly pronounced diel rhythms, likely as a response to heterogenous habitat types along with temporal and spatial variations in water chemistry (Coe, 1966; Coe, 1967). Coe (1967) measured large temperature changes in Lake Magadi lagoons, with the shallow waters rapidly dropping in temperature, from up to 36°C during the middle of the day down to 20°C after sunset. This temperature change corresponded with mass migration to warmer and deeper parts of the lagoon or closer to the warm spring inlets (Coe, 1967). Albrecht noted that male *A. alcalica* would feed on algae early in the morning before moving to the breeding pits, with courtship and spawning occurring more frequently after midday. Conversely, *A. grahami* males would feed at night, with greater sexual activity occurring before midday (Albrecht, 1967).

1.3.7 Geology and biogeography of the Magadi-Natron basin

The Magadi-Natron basin is situated ~605 m above sea level, and both lakes have a similar structure and are described as seasonally flooded, volcanic and endorheic basins with alkaline hydrothermal springs (Williamson *et al.*, 1993; Owen *et al.*, 2019). The interior of the lakes consists of sodium hydrogencarbonate precipitates (trona), with discrete shallow-water lagoons concentrated around the lake margins (Kaufman *et al.*, 1990; Seegers and Tichy, 1999; Pörtner *et al.*, 2010). The lakes experience significant fluctuations in water levels between the dry and rainy seasons. In the dry season, most of the central part of the lake consists of soda mud flats (Tebbs *et al.*, 2013). Lake Natron and Magadi are situated ~36km apart and are currently not connected by rivers or streams (Nagl *et al.*, 2001). Taken from satellite images, Lake Natron has an average surface area of 398 km^2, varying from 81-804 km^2 depending on the dry/wet seasons (Tebbs *et al.*, 2013). Magadi has a surface area of 75 km^2 in the dry season to 108 km^2 in the wet season (Jones, 1977).

The high levels of trona are a result of the negative evaporative balance of the Natron-Magadi basin (Burrough and Thomas, 2009). In Natron, 27% of the lake's water originates from the Ewaso Ngiro South River (which enters the lake from the north), 25% comes from the saline springs, 39% comes from direct rainfall on the surface of the lake, with the remaining 9% originating from other rivers in the catchment (Tebbs *et al.*, 2013). Conversely, Lake Magadi has no obvious inflow from rivers, and the

lake's water is almost exclusively derived from rainfall and volcanic springs (Butzer *et al.*, 1972).

Lake Natron and Magadi are the remnants of a large paleaolake called Orolonga that formed approximately 700 ka. The wider basin has existed since the middle Pleistocene from 1.7 Ma (Eugster, 1986). The basin likely underwent multiple humid events during interglacial periods where interspersed lakes were connected by rivers (Deino *et al.*, 2019). The lake then contracted in size during the intense arid event called the Younger Dryas ~13 to 12 ka (Williamson *et al.*, 1993; Owen *et al.*, 2018; Deino *et al.*, 2019), but Natron and Magadi formed a single lake in the humid period from ~12.4 to 8.2 ka (Dommain *et al.*, 2022). At its highest extent, the lake is estimated to have had a maximum depth of around 50-60m (Roberts *et al.*, 1993; Dommain *et al.*, 2022). Also, during this period, the joint Magadi-Natron lake was intermittently connected to palaeolakes Siriata and Naivasha, forming a network of lakes connected by the Southern River (Dommain *et al.*, 2022). Approximately 8 ka, the lake had split into Lakes Natron, Magadi and Little Magadi (Butzer *et al.*, 1972; Williamson *et al.*, 1993; Dommain *et al.*, 2022) with increasing contraction occurring since then (Roberts, 1993; Seegers *et al.*, 2001). The hypersaline and hyperalkaline conditions are thought to have emerged soon after this period (~7 ka) (Butzer *et al.*, 1972; Roberts *et al.*, 1993; Tichy and Seegers, 1999).

Fossil cichlids have been recovered from deposits surrounding Lake Magadi. About 12m above the present lake level are the High Magadi Beds, where *Alcolapia*-like fossils have been found, although they are of a much larger size (~200mm SL) (Coe, 1966). These fossils were later calibrated to an age of 7070±180 years, which corresponds to when the lakes were deeper, but still had slightly saline-alkaline conditions (Butzer *et al.*, 1972; Roberts *et al.*, 1993; Dommain *et al.*, 2022). Other collections around the northwest and southeast of Lake Magadi uncovered fossil fish of a smaller length (30-100mm), which were dated to 7950±210 years, corresponding to the single merged lake (Tichy and Seegers, 1999; Dommain *et al.*, 2022). In addition, a fossil scale resembling *Alcolapia* was found in Palaeoake Siriata, a basin ~11 km from Magadi and was dated from 11.4±0.6 ka (Dommain *et al.*, 2022). These larger fish fossil findings originating from the deeper, freshwater palaeolake suggests

that *Alcolapia* were more widespread and potentially only had generalist trophic morphology, with the more derived downturned mouths developing more recently.

Much older and more distant *Alcolapia*-like fossils have been recorded ~285 km to the north of Magadi in the Tugen Hills, Kenya. These fossils were found to share more features with *Alcolapia* than other extant *Oreochromis,* yet existed in alkaline palaeolakes during the early Miocene 10-13 Ma (Penk *et al.*, 2019; Kevrekidis *et al.*, 2020). Such fossils dating from this far back in the past suggest that soda lake adaptations may have arisen multiple times during climatic shifts.

1.4 Thesis aims

The recent, small-scale and isolated nature of the *Alcolapia* radiation makes it ideally suited for studying the various processes that have led to the formation of this radiation. Previous studies identified distinct morphological diversity yet low genomic differentiation and high levels of admixture (Ford *et al.* 2015; Kavemebe *et al.* 2016), raising many questions about the factors governing their divergence. In this thesis, we employed an integrative approach to study the forces influencing *Alcolapia* speciation. We explored the role of premating barriers in maintaining reproductive isolation, investigated the degree of ecological partitioning between species, and inferred the evolutionary history of this radiation.

In Chapter 2, we characterised the degree of mate choice between sympatric and allopatric *Alcolapia* species using laboratory mating trials to:

1. Determine the extent to which *Alcolapia* exhibit assortative mating.
2. Test if sympatric Natron *Alcolapia* species show evidence of stronger assortative mating compared to the allopatric species *A. grahami.*

In Chapter 3, we conducted the molecular characterisation of the Natron *Alcolapia* diet and gut microbiota to:

1. Assess the extent to which Natron *Alcolapia* gut microbiota differ from other cichlid radiations, including the presence of extremophile taxa.

2. Test if sympatric *Alcolapia* species exhibit cryptic differences in diet and gut microbiota representing ecological resource partitioning.

In Chapter 4, we used genomic datasets to conduct population genetic analysis and utilised demographic modelling techniques to:

1. Calculate the relative rates of gene flow between *Alcolapia* populations.

2. Infer changes in population size, divergence times, and contemporary and historical gene flow in the *Alcolapia* radiation.

Finally, in Chapter 5, we discuss the context of the findings within the wider literature and outline future directions for research.

CHAPTER 2. SYMPATRIC AND ALLOPATRIC SODA LAKE CICHLID SPECIES SHOW SIMILAR LEVELS OF ASSORTATIVE MATING

ABSTRACT

Characterising reproductive barriers such as mating preferences within rapid evolutionary radiations is crucial for understanding the early stages of speciation. Cichlid fishes are well-known for their adaptive radiations and capacity for rapid speciation, and as such, we investigate assortative mating among *Alcolapia* species; a recent (<10,000 years), small adaptive radiation, endemic to the extreme soda lakes, Magadi (one species) and Natron (three species), in East Africa. In seminatural aquarium conditions, we observed both courtship and mate choice, tested by microsatellite paternity analysis, to be significantly assortative among the three sympatric Natron species in a three-way choice experiment. This was also the case between allopatric species from Natron and Magadi, as found in a two-way choice experiment. The proportion of disassortative matings was non-trivial in both of these experiments, with hybrids comprising 29% of offspring in sympatric species and 11.4% in allopatric species comparisons. Previous work suggests that the Natron/Magadi split might not be much older than the radiation within Natron. Hence, the relatively lower rate of hybridisation in the allopatric comparison is surprising and inconsistent with predictions of reinforcement theory, which predicts a faster rate of accumulation of premating isolation in sympatry. The relatively weak assortative mating in sympatry suggests that additional reproductive barriers, such as microhabitat preferences or spatial structuring, may contribute to genetic isolation in nature.

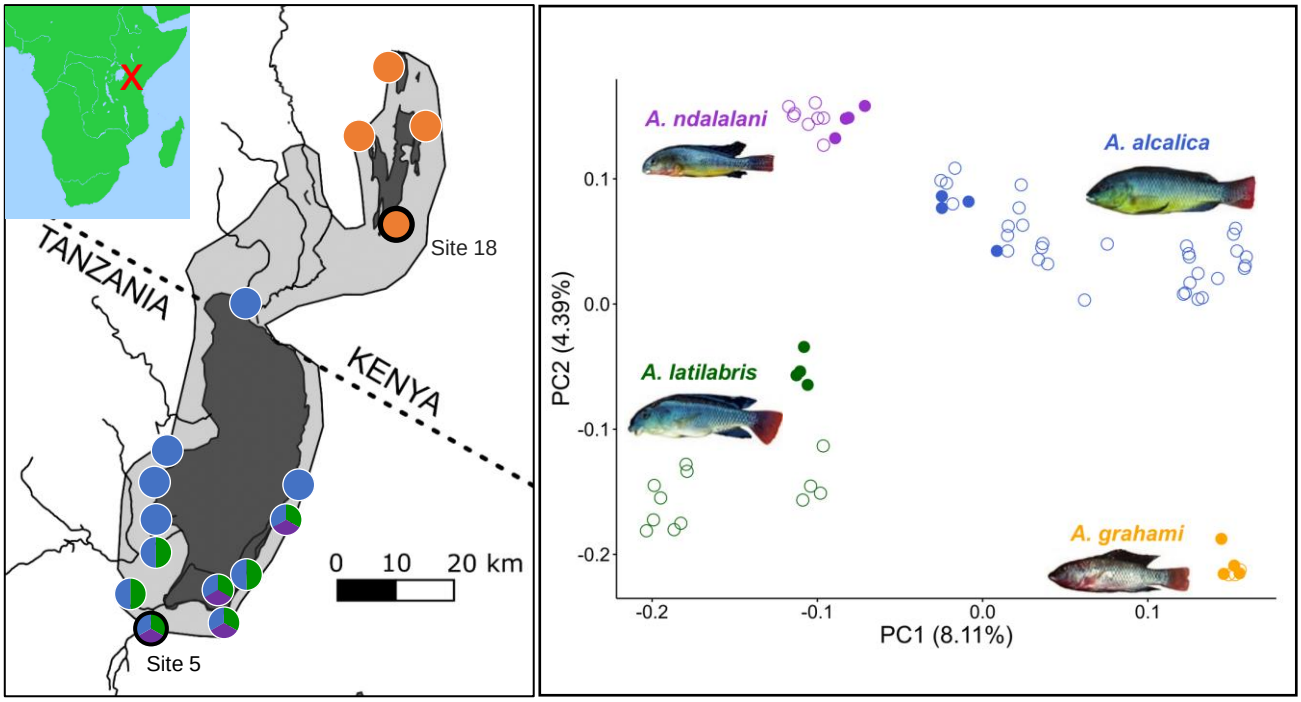


Figure 2.1. Distribution and species relationships of the *Alcolapia* spp. radiation of lakes Natron and Magadi. Left panel: Map showing the distribution of the species from sites that have been sampled. The light grey area shows the approximate maximum extent of Palaeolake Orolonga (from ~700 Ka) (Williamson *et al.,* 1993). The dark grey area represents the current expanse of the lakes. Coloured circles correspond to species present in the right panel with black-ringed circles denoting the populations used in this study. Photos and site information are from Ford *et al.,* 2015. Right panel: Principal Component Analysis (PCA) of a published *Alcolapia* RADseq dataset (Ford *et al.,* 2015) demonstrates that the species form distinct genetic clusters even in sympatry. Colours correspond to species in the left-hand panel. Filled circles represent the populations used in this study (site 5 for sympatric species and site 18 for *A. grahami*, the allopatric species); empty circles are individuals from other sites.

2.1 Introduction

Speciation can be best understood by studying the emergence of reproductive barriers within a previously interbreeding population (Coyne and Orr, 2004). As speciation progresses, gene flow between the diverging taxa diminishes as a consequence of the strengthening of existing barriers, and/or the accumulation of other barriers, eventually leading to complete reproductive isolation (Kulmuni *et al.,* 2020). In sexually reproducing organisms, reproductive barriers can be prezygotic or postzygotic. Postzygotic barriers can be extrinsic, and associated with adaptation to divergent

environments, or intrinsic and dependent upon genetic incompatibilities that occur irrespective of ecology (Coyne and Orr, 2004). The degree of reproductive isolation changes as speciation proceeds, with the order and appearance of reproductive barriers varying at different stages of the speciation continuum (Drès and Mallet, 2002; Stankowski and Ravinet, 2021). Therefore, characterising reproductive barriers at different stages along this continuum is important for creating a complete picture of speciation, and mechanisms initiating speciation are best studied by focusing on taxa at the earliest stages along the continuum (Coyne and Orr, 1997; Butlin *et al.,* 2012).

Geographic context can have a major influence on the speciation process as it affects the levels of gene flow between diverging populations (Coyne and Orr, 1997; Rosser and Queste *et al.,* 2019). Classical models of speciation focused on the role of geographic barriers in the formation of new species, where populations become physically separated and diverge under the effects of local adaptation and genetic drift (Mayr, 1947; Turelli *et al.,* 2001). Sympatric or parapatric speciation, which occurs in the absence of physical barriers and in the presence of gene flow, generally arises when premating isolation becomes associated with a trait under divergent selection (Coyne and Orr, 2004; Smadja and Butlin, 2011, Servedio *et al.,* 2011). Reproductive character displacement, where there is a greater divergence of reproductive traits in sympatry compared to allopatry, can occur due to reinforcement, whereby selection strengthens premating barriers that reduce hybridisation rates (Dobzhansky, 1940; Pfennig and Pfennig, 2009), or via reproductive interference among fully isolated species (Templeton, 1981). However, premating isolation can emerge in allopatry if adaptation to different ecological pressures is associated with reproductive barrier traits (Servedio *et al.,* 2011), as a by-product of sexual selection (Fisher, 1930; Mendelson and Safran, 2021), or through mutation-order effects (Mani and Clarke, 1990). Therefore, to provide evidence of the mechanisms driving speciation under gene flow, it can be informative to include contrasts with recently separated allopatric populations (Coyne and Orr, 1997; Funk, 1998).

Cichlid fishes are well known for their striking adaptive radiations in the East African Great Lakes, where hundreds of species have evolved from a single or handful of closely related ancestral species often over very short timescales (Turner *et al.,* 2001; Salzburger and Meyer, 2004). Many mechanisms drive speciation in cichlids, including

ecologically-mediated processes such as diet or habitat depth (Albertson *et al.,* 2003; Terai *et al.,* 2006) and sexually-mediated processes such as assortative mate choice (e.g. Knight *et al.,* 1998, Seehausen *et al.,* 1998). Further factors influencing diversification include introgressive hybridisation (Salzburger *et al.,* 2002), the reassembly of old genetic variants into new combinations (Meier *et al.,* 2017a; Marques *et al.,* 2019) or geographic isolation (Sturmbauer *et al.,* 2001). Additionally, multiple processes may operate together, an example being sensory drive (Seehausen *et al.,* 2008). Between sympatric cichlid species, isolating mechanisms are more often prezygotic, with female choice generally being the ultimate barrier to mating (Kocher, 2004; Henning and Meyer, 2014).

Understanding the emergence of reproductive barriers in larger and/or older systems such as the East African Great Lakes is often difficult due to the complications of historical lake level fluctuations combined with the complex evolutionary history of ancestral lineages, with past periods of gene flow and extensive incomplete lineage sorting making inferences more difficult (Malinksy *et al.,* 2015; Svardal *et al.,* 2021). Therefore, recent and smaller radiations from isolated lakes make for more tractable study systems, where it may be possible to disentangle both the order of emergence and relative contribution of different reproductive barriers (Barluenga *et al.,* 2006; Malinsky *et al.,* 2015; Poelstra *et al.,* 2018; Kautt *et al.,* 2018). Direct tests of the levels of assortative mating in such simple systems have thus far only been reported between Midas cichlids from Nicaraguan crater lakes (Elmer *et al.,* 2009; Machado-Schiaffino *et al.,* 2017).

In this study, we investigated the magnitude of assortative mate choice among the four closely related *Alcolapia* species which comprise a young, and isolated adaptive radiation endemic to the East African soda Lakes Magadi, in Kenya, and Natron, in Tanzania (Seegers and Tichy, 1999). Although widely referred to as a separate genus (Kavembe *et al.,* 2014; White *et al.,* 2020), the *Alcolapia* clade is in fact nested within the genus *Oreochromis* (Ford *et al.,* 2019). Magadi and Natron are volcanic, alkaline lakes dominated by large areas of thick sodium hydrogencarbonate precipitates, with very shallow (< 1 m) lagoons, streams and hot springs interspersed around the lake margins (Kaufman *et al.,* 1990; Seegers *et al.*, 1999). *Alcolapia* are the only fishes found in these lakes and have evolved several unique adaptations to thrive in

extremes of pH, temperature, salinity, UV light and oxygen levels (Trewavas, 1983; Narahara *et al.,* 1996; Walsh *et al.,* 2001; Wood *et al.*, 2012, 2016; White *et al.,* 2020). The much larger and deeper (~ 55 m) paleolake Orolonga, which comprised part of rift lake network intermittently connected by rivers, contracted and split to form Natron and Magadi ~ 8 Ka. Orolonga itself had more freshwater conditions, and the current highly alkaline and hypersaline conditions are thought to have developed ~ 7 Ka (Roberts *et al.,* 1993; Dommain *et al.,* 2022). Therefore, both the adaptive evolution and the speciation of *Alcolapia* has been extremely rapid.

Three Alcolapia species are described from Lake Natron, *Alcolapia alcalica, A. ndalalani and A. latilabris*. A single species, *A. grahami*, is known from Lakes Magadi and Little Magadi (Coe, 1966; Seegers *et al.,* 2001). The species differ in morphology, size and male nuptial colouration (Seegers and Tichy, 2001; Figure 2.1). The Natron species have different head and mouth shapes which likely relate to fine-scale niche specialisation towards different forms of herbivory (Ford *et al.,* 2016). *Alcolapia* populations are distributed across the springs, lagoons and small streams around the perimeters of Natron and Magadi. While the distribution of the three species around Lake Natron is uneven, there are sites in the south of the lake where all three species can be found swimming alongside each other at high densities (Ford *et al.,* 2015; Figure 2.1). Genomic data revealed evidence of ongoing gene flow between all the sympatric Natron species and extremely low genomic differentiation between species (Ford *et al.,* 2015), even when compared to other cichlid radiations (Svardal *et al.,* 2021). Despite this gene flow, these species are genetically distinct (Figure 2.1).

In *Alcolapia*, high population density, conspicuous male nuptial colouration, presence of leks and male-biased sex ratios are all predictors of high levels of sexual selection (Seegers and Tichy, 1999; Ford *et al.,* 2016; Maina *et al.,* 2019). Differences in male colouration between the three sympatric Natron species suggest that prezygotic assortative mate choice could be an important factor in their reproductive isolation. Since a comparison between sympatric and allopatric taxa can be valuable in understanding the mechanisms driving speciation with gene flow, in this chapter we used semi-natural aquarium conditions to quantify the levels of assortative mate choice (male courtship behaviour and paternity of embryos) both among sympatric Natron *Alcolapia* species (*A. alcalica*, *A. latilabris* and *A. ndalalani*) and also between

allopatric species (Natron *A. alcalica* and Magadi *A. grahami*). We used this data to test the hypotheses that sympatric *Alcolapia* exhibit stronger assortative mate compared to the allopatric species *A. grahami.*

2.2 Material and Methods

2.2.1 Fish collection and husbandry

The three Natron species were collected from site 5 in July 2017 (Ford *et al.,* 2015; White *et al.,* 2020) (COSTECH permit no. 2017-259-NA-2011-182), and *A. grahami* collected from Lake Magadi site 18 (Ford *et al.,* 2015) in March 2019 (NACOSTI research permit no. NACOSTI/P/18/79668/24717) (Figure 2.1). Fishes were kept in separate single-sex stock tanks at Bangor University. Appropriate water chemistry was maintained (pH = 9.0, GH = 180ppm, KH = 180ppm, specific gravity = 1.005) by the addition of $NaHCO_3$ (0.5 g/L), Na_2CO_3 (0.07 g/L), $MgSO_4.7H_2O$ (1.5 g/L), Instant Ocean® Sea Salt (0.5 g/L)), and the tanks were kept at 31 °C under a 12:12 light:dark cycle. Fish were fed daily with spirulina flake food. Euthanasia was carried out according to the schedule 1 protocol for humane killing (as indicated in Animals Scientific Procedures Act 1986). Work was approved after an ethical review by the University of York and Bangor University.

2.2.2 Experimental setup

To quantify the degree of assortative mating in *Alcolapia* and to determine whether this is affected by geographic context, we carried out two separate mate choice experiments in aquarium setups; among species that are found in sympatry, and between species that have allopatric distributions. The strength of assortative mating was assessed through observations of courtship behaviour and paternity analysis of resulting offspring.

The sympatric mate choice experiment involving the Natron species (*A. alcalica*, *A. latilabris* and *A. ndalalani*) was carried out between February and March 2019. For the allopatric experiment (*A. alcalica* and *A. grahami*), behavioural measurements were carried out between January and February 2020, while broods were collected between March and April 2020. The sympatric mate choice experiment was carried out in a single 8 m (L) x 0.6 m (W) x 0.7 m (D) tank. The allopatric mate choice experiment

was carried out in the same tank, but reduced to a length of 6 m, to maintain approximately the same density of fish. An 8 cm layer of silica sand was used as a substrate, with shelter provided by evenly placing five clay pots for every 2 m of tank length.

Fully mature males were selected for both experiments and size-matched with a range in standard length (SL) of no more than 12 mm within species. For the sympatric experiment, fish were all first-generation laboratory-bred and consisted of 10 females and 6 males of each of the Natron species derived from the wild-caught parents. In the sympatric experiment, males had the following mean SL±SD: *A. alcalica* 76.2±4.9 mm, *A. latilabris* 68.7±4.0 mm, *A. ndalalani* 65.8±3.0 mm. The larger SL of *A. alcalica* males relative to the other sympatric species reflects inherent size differences of the species observed in the wild (Seegers and Tichy, 1999). In the allopatric experiment, fish consisted of 12 females and 7 males each of first-generation laboratory-bred *A. alcalica* and wild-caught *A. grahami*. *A*. *alcalica* was selected as the allopatric Natron species because it is the most widely distributed and forms the basal lineage (Ford *et al.,* 2015). *A. alcalica* also possesses similar trophic morphology to *A. grahami,* with both sharing a terminal mouth (Seegers and Tichy, 1999; Ford *et al.,* 2016). Mean male SL was 72.1±3.7 mm for *A. alcalica* and 71.1±4.0 mm for *A. grahami*.

For both experiments, males and females were kept in single-sex stock tanks for at least a month before being added simultaneously to the experimental setup. While males of all the *Alcolapia* species and female Natron species are easily differentiated by their colouration and unique mouth morphology, female *A. grahami* and *A. alcalica* are difficult to differentiate. Therefore, the females in the allopatric experiment were made visually distinguishable by caudal fin clips of a single species at a time. The females were fin-clipped every two weeks after the fin section had almost re-grown, and the species of clipped fish was alternated.

2.2.3 Courtship behaviour measurements

Identification of focal males was possible using reference photos of their unique scale markings along the intersection of the tail and caudal fin. Individual identification was not possible for females or for male opponents with whom interactions were too rapid. Instead, the species, sex and relative size of the interacting fish was recorded.

Table 2.1. Description of behavioural patterns scored during focal male observations. Derived from Baerends and Baerends van Roon (1950). PE – point event, SE – state event.

Behavioural pattern		**Description**
Courtship (any behaviours)	SE	Any courtship behaviours performed by the male.
Tilting	PE	Fish leans horizontally, presenting flank towards female. Dorsal and anal fin close to body.
Leading	PE	While tilting fish swims in front of female towards bower.
Circling	PE	Male and female circle the nest together.
Quivering	PE	Male rapidly vibrates body to release sperm.

Behaviours of randomly selected individual males were scored during daily, 5-minute focal observations between 10:00 and 13:00 over a period of 15 days.

Since many of the courtship behaviours in *Alcolapia* were found to be similar to those described for other *Oreochromis* species (e.g. tilting, circling and quivering), a reduced ethogram of behaviours was created based on the descriptions by Baerends and Baerends van Roon (1950; Table 2.1). Courtship was scored and its duration was recorded if the male performed any courtship behaviours directed at a female within a distance of two SL of the focal male. Courtship directed towards multiple females of different species was recorded but not assigned a species. For the allopatric experiment, focal observations were video recorded and were carried out blind with respect to the identity of the female species (females were scored as either fin-clipped or non-fin-clipped). Courtship behaviour was then scored using Solomon Coder v. 19.08.02 (Péter, 2011) using a custom ethogram.

2.2.4 Paternity testing

Alcolapia are maternal mouthbrooders and females were checked each day visually for brooding. The partially developed broods were removed from females, euthanised and counted. Each time a brood was removed from a female, a sample of the female's DNA was obtained by swabbing the fish along the body using sterile cotton swabs (Breacker *et al.,* 2017). All males were photographed, measured (SL) and swabbed for DNA before being placed into the setup. Swabs and embryos were stored in 95%

ethanol at -20 °C until required for DNA extraction. In the sympatric experiment, 12-13 separate broods were collected from each species during the same period when courtship measurements were taken. While 15 separate broods were collected per species in the allopatric experiment, these were collected after the period during which the courtship measurements were made.

Swab DNA was extracted following the protocol outlined in Breacker *et al.* (2017). To assess paternity, DNA was extracted from 2-3 mm sections of tissue from up to six randomly selected embryos per brood. For most embryo DNA extractions, a modified version of the swab protocol was used by replacing the swab with dissected tissue. For a few embryos, a DNeasy Blood and Tissue Kit (Qiagen Inc.) was used. Extracted DNA was diluted 1:5 using ddH_2O for use in PCR.

Dinucleotide microsatellites were detected in the *Oreochromis niloticus* genome (O_niloticus_UMD_NMBU; Conte *et al.,* 2017) using SciRoKo (Kofler *et al.,* 2007), after which 2-3 primers pairs per chromosome were designed using Primer-BLAST (Ye *et al.,* 2012). Initial screening of 53 dinucleotide-repeat microsatellites for polymorphism was carried out using a panel of 8 individuals of the three Natron species using FAM-labelled M13-tailed primers (Table 2.2) (Schuelke, 2000). Seven loci exhibiting within and among species polymorphism were selected for use in paternity testing. Microsatellite DNA loci were amplified in two separate multiplexes of fluorescent-tagged primers using the Type-IT Microsatellite PCR Kits (Qiagen) in 10 µl PCR reaction volumes using the manufacturer's standard protocol (Table 2.2). The PCRs were run with an initial 95 °C denaturation stage for 5 min before conducting 32 cycles: 95 °C for 30 s, 60 °C for 90 s and 72 °C for 30 s, and a final stage of annealing at 60 °C for 30 min. PCR products were diluted 1:5 before being analysed on an ABI 3730xl DNA Analyzer (Thermo Fisher, Massachusetts, USA). Allele sizes were scored automatically using GeneMarker v. 2.6.2 (SoftGenetics, LLC., Pennsylvania, USA). All traces were manually inspected and corrected where necessary to ensure high genotype data quality.

Microsatellite genotypes were manually scored using *Microsatellite Analysis Software* (*MSA*) (Thermo Fisher, Massachusetts, USA). First, to identify which swabbed

Table 2.2. PCR Primers used for microsatellite paternity analysis of *Alcolapia* offspring. Two sets of fluorescent-tagged primers (1 and 2) were run in separate multiplexed PCRs.

PCR multiplex mix	Primer name	Forward (5'-3')	Reverse (5'-3')	Fluorescent dye
1	967.3	CTG TTT CTT TGC CCA AAA CGG T	CCA ATG ACC GTG CTT ACA GGA	NED
	974.3	TGG AGG ATG CGA CGC TCA TTT	CTG TGA AGC GTT TTT CTG GGG TA	VIC
	981.1	ACA CAG ACG GTC ATA ATC CCT G	GGA ACA GTT TAA GGC ACA AGT CC	PET
	983.3	TGT ACT AAG CAG CTC CCA GGT	GGT GGA AAT ACG CAC AAC GA	6-FAM™
2	978.3	AAG CGT ACA TAA GGC CAG TG	GAC TCT CTA TCC ATC CGA CCG	6-FAM™
	981.3	GGG GGC AAT GCC ATT AAA AAG A	GGC TTG GAA TTG TGA GAG GAA	PET
	986.1	TTC ACC TCG GCC ATT AAT CCT T	TAG GCA GGC GAA CTC TCA CC	VIC

females had matching genotypes, an identity analysis was carried out using *cervus* v.3.0.7 (Kalinowski *et al.,* 2007) using all 7 loci (Table 2.2). In addition, sequencing

failed for one female (*A. ndalalani* 10) and therefore to check if this genotype matched with any of the other females, genotype reconstruction was performed in *colony2* using its known offspring and all candidate males (Jones and Wang, 2010). Parentage analysis was carried out for both the sympatric and allopatric experiments separately using *cervus*. Allele frequencies were generated using all the parental genotypes (Flanagan and Jones, 2019), and a simulation of paternity was run with 100,000 offspring using all possible candidate fathers. A threshold minimum of 4 typed loci was used for paternity assignment. Paternity was assigned based on the LOD score of offspring and parent trios. The trio LOD score is calculated using the genotypes of offspring, known mothers and candidate males while accounting for potential mistyping errors. First, all offspring were assigned a single compatible male if the trio LOD score had assignment confidence of at least 95%. For the remaining offspring, candidate fathers were removed if they had more than one mismatch in either pairwise (offspring-candidate father) or trio comparisons. For the single female with no genotype information, only candidate fathers with a positive pairwise LOD score were kept.

2.2.5 Statistical analysis

Courtship preference towards different taxa was modelled using Generalized Linear Mixed Models (GLMMs) using two different response metrics: 1) the total amount of time spent courting and 2) courtship frequency, or the total number of courtship behaviours directed towards females. Both response measures were the sum of behaviours carried out over a 5-minute focal observation. For both models, fixed effects included the species of male and species of female involved in courtship, whereas individual male ID and date of observation were modelled as random factors, correcting for pseudoreplication. In addition, to account for potential temporal variations in courtship, time of observation was modelled as an additional random effect with times of day split into 5-minute intervals. Models with and without the time of day were selected depending on the Akaike Information Criterion (AIC). Courtship time and courtship frequency behaviours were modelled with a separate model for each experiment. For the courtship time models, the response had a heavily right-skewed distribution with many zeros, therefore, models used a zero-inflated gamma distribution with a log link. We allowed zero-inflation to vary within each level of the fixed effects. As courtship frequency consists of count data, these models used a Poisson distribution and log link, but due to overdispersion in the allopatric experiment, a negative binomial distribution (*nbinom2*) with a log link was used instead.

Assortative mate preferences were also tested using the offspring paternity data. Broods were aggregated for each individual female and offspring were then scored as being either of conspecific or heterospecific paternity. Mating preference towards conspecifics was modelled using a Generalized Linear Model (GLM) with the *cbind* function and a beta-binomial distribution to account for overdispersion. Due to insufficient data, it was not possible to account for variance among individuals by using a mixed model design. The proportion of conspecific broods was modelled as the response variable and the species of the mother as the independent variable. In addition, differences in total brood size between species were tested using a GLMM with a Poisson distribution and individual female ID modelled as a random effect.

For all models, estimates and post-hoc contrasts were generated using the *emmeans* package (Lenth *et al.,* 2018). Model estimates were obtained for each species of their overall courtship propensity towards any species and towards different species. The

significance (*p*-values) of differences in courtship between taxa was obtained through pairwise comparisons, with Tukey adjustments to account for multiple contrasts. To obtain estimates of the degree of assortative mating from offspring paternity data, the predicted probabilities of mating with a conspecific male were obtained from the output of the GLM. To test for the significance of assortative mate choice for each taxon, *p*-values were obtained by testing if the predicted probability of mating with conspecifics was significantly different from the expected proportion of conspecific matings under random mating: 0.5 in the allopatric experiment (two-way choice) and 0.33 in the sympatric experiment (three-way choice).

All statistical analyses were carried out using R version 4.1.2 (R Core Team, 2021). The packages *lme4* (Bates *et al.,* 2014) and *glmmTMB* (Brooks *et al.,* 2017) were used to generate models (Table A2.1). *DHARMa* (Hartig, 2021) was used to test model assumptions and the fit of each model. Full model commands and output are provided in Table A2.1. To visualise parental-offspring relationships, a network was created with the R package *tidygraph* v1.2.1 (Pedersen, 2022a) using additional code maintained by James Ward (https://jmw86069.github.io/multienrichjam/) to create a node layout using the Fruchterman-Reingold algorithm. The package *ggraph* v.2.0.5 (Pedersen, 2022b) was used to plot networks while *ggplot2* v3.3.6 was used to generate all other plots (Wickham, 2016).

2.3 Results

2.3.1 Observations of territoriality and courtship in *Alcolapia*

Males were highly active and performed courtship and territorial behaviours soon after their introduction to the experimental setup. Between courtship and aggressive behaviours, dominant males spent a significant amount of time constructing bowers, simulating the lekking areas found in the wild (Coe, 1969). While individual males often held the same bower for multiple days, bower ownership also changed frequently over the course of the experiments. Males directed the most aggressive behaviours towards males that approached their territory or interrupted courtship and mating. Aggressive behaviours were directed both toward conspecific and heterospecific males, with the larger *A. alcalica* males tending to be the most successful in winning

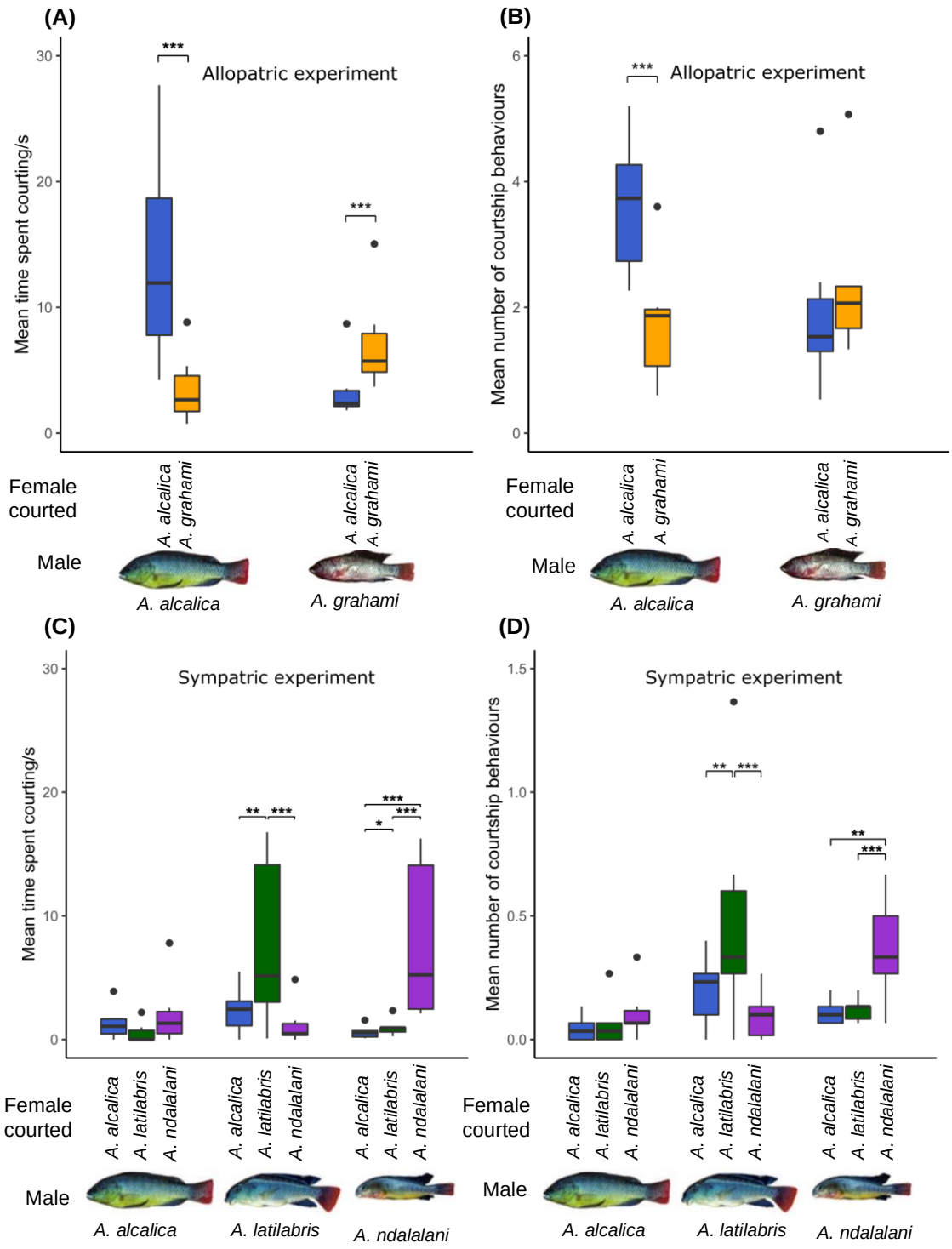


Figure 2.2. Differences in male courtship with conspecific and heterospecific females in *Alcolapia*. Recorded are (**A,C**) time spent courting and, (**B,D**) number of courtship attempts for each male during 5 minute focal observations. Courtship results are shown for both the allopatric (**A,B**) and sympatric experiments (**C,D**). Colours denote the species of female courted with: blue = *A. alcalica*, green = *A. latilabris*, purple = *A. ndalalani*, orange = *A. grahami*. Large outliers (35.4s and 57.6s) were removed from panel (**C**) for clarity (see Figure A2.1 for full figure). Asterisks denote significant levels of pairwise contrasts extracted from GLMMs: *** < 0.001, ** < 0.01, * < 0.05.

agonistic encounters in the sympatric experiment. Subdominant males were characterised by their less bright colouration, lower levels of courtship, lack of aggression towards other males and lack of bower ownership. Subdominant males were also observed to shoal with groups of females and avoid the bowers of other males. Aggression was infrequent amongst females.

Courtship was common, but spawning was observed less frequently. While some females were observed to mate with the same male on multiple occasions, other females were observed to spawn with several males within a single brood. In a small number of cases, potential sneak mating was observed, where spawning was interrupted by a rival male as the female released an egg. In sympatric species, brood size ranged from 5-31 with a mean (±SD) of 20.1±7.9 *A. alcalica*, 9.8±4.4 *A. latilabris* and 15.7±6.6 *A. ndalalani*. *A. latilabris* had significantly smaller brood sizes than both *A. alcalica* (GLMM, post-hoc, $p < 0.001$) and *A. ndalalani* ($p < 0.006$; Table A2.2). However, comparisons between *A. alcalica* and *A. ndalalani* were not significant (GLMM, post-hoc, $p < 0.14$; Table A2.2). In the allopatric experiment, brood size ranged from 3-51 and *A. alcalica* had a larger mean (±SD) brood size of 28.2±12.6 compared to 19.1±13.4 in *A. grahami*, but the difference was marginally non-significant (GLMM, post-hoc, $p < 0.064$; Table A2.2).

2.3.2 Assortative male courtship behaviour

Male courtship behaviours are based on 1050 and 1350 minutes of observation data for the allopatric and sympatric experiments respectively. In the allopatric experiment, males of both species spent significantly more time courting conspecific than heterospecific females (GLMM post-hoc, *A. alcalica*: $p < 0.001$; *A. grahami*: $p < 0.001$; Figure 2.2A; Table A2.3). Additionally, overall courtship time with any species did not differ between males or females of each species (males: $p = 0.54$, females: $p = 0.65$; Table A2.3). Contrastingly, *A. alcalica* males had a higher courtship frequency (number of courtship attempts) with conspecific females, while *A. grahami* males did not (*A. alcalica*: $p < 0.001$, *A. grahami*: $p = 0.18$; Figure 2.2C; Table A2.4). This was despite there being no differences in overall courtship frequency (courtship towards any species) between males of both species ($p = 0.27$; Table A2.4). Likewise, *A. alcalica* females were courted by conspecifics significantly more often, while *A. grahami* females were not (*A. alcalica*: $p = 0.002$, *A. grahami* $p = 0.23$; Table A2.4);

however *A. alcalica* females also had a higher overall courtship frequency compared to *A. grahami* females ($p = 0.015$; Table A2.4).

In the sympatric experiment, *A. latilabris* males spent significantly longer courting *A. latilabris* females than *A. alcalica* females ($p = 0.001$) and *A. ndalalani* females ($p < 0.001$), but there was no significant difference in time spent courting between both heterospecific species ($p = 0.32$; Figure 2.2A; Table A2.6). *A. ndalalani* males spent significantly longer courting conspecific females compared to *A. alcalica* females ($p < 0.001$) and *A. latilabris* females ($p < 0.001$), but there was no difference in courtship time in the two heterospecific comparisons ($p = 0.21$; Figure 2.2A). In contrast, there was no significant difference in the amount of time spent courting between any of the species by *A. alcalica* males (*A. alcalica* - *A. latilabris*: $p = 0.96$, *A. alcalica* - *A. ndalalani*: $p = 0.73$, *A. ndalalani* - *A. latilabris*: $p = 0.52$; Figure 2.2A; Table A2.6). There were no significant differences in overall courtship time between males (*A. alcalica* - *A. latilabris*: $p = 0.99$, *A. alcalica* - *A. ndalalani*: $p = 0.63$, *A. ndalalani* - *A. latilabris*: $p = 0.57$; Table). However, *A. ndalalani* females courted for significantly longer overall compared to *A. alcalica* females ($p = 0.018$; Table A2.6).

The number of courtship events largely reflected these results with *A. latilabris* males courting conspecific females significantly more often than heterospecifics (*A. latilabris* - *A. alcalica*: $p = 0.003$, *A. latilabris* - *A. ndalalani*: $p < 0.001$, *A. ndalalani* - *A. alcalica*: $p = 0.18$; Figure 2.2B; Table A2.8). Similarly, *A. ndalalani* males also spent had a higher courtship frequency with conspecifics compared to heterospecifics (*A. ndalalani* - *A. alcalica*: $p = 0.002$, *A. ndalalani* - *A. latilabris*: $p = 0.003$, *A. latilabris* - *A. alcalica*: $p = 0.97$; Figure 2.2B; Table A2.8). There was no significant difference in courtship frequency between any of the species by *A. alcalica* males (*A. alcalica* - *A. latilabris*: $p = 0.79$, *A. alcalica* - *A. ndalalani*: $p = 0.24$, *A. ndalalani* - *A. latilabris*: $p = 0.56$; Figure 2.2B; Table A2.8). There were no significant differences in overall courtship frequency between males of each species (*A. alcalica* - *A. latilabris*: $p = 0.19$, *A. alcalica* - *A. ndalalani*: $p = 0.16$, *A. ndalalani* - *A. latilabris*: $p = 0.99$). However, *A. alcalica* females had an overall lower courtship frequency compared to the other species (*A. alcalica* – *A. latilabris*: $p = 0.01$, *A. alcalica* – *A. ndalalani*: $p = 0.03$, *A. latilabris* – *A. ndalalani*: $p = 0.91$; Table A2.8). Full model outputs and associated test statistics can be found in supplementary Tables (Table A2.1-A2.8).

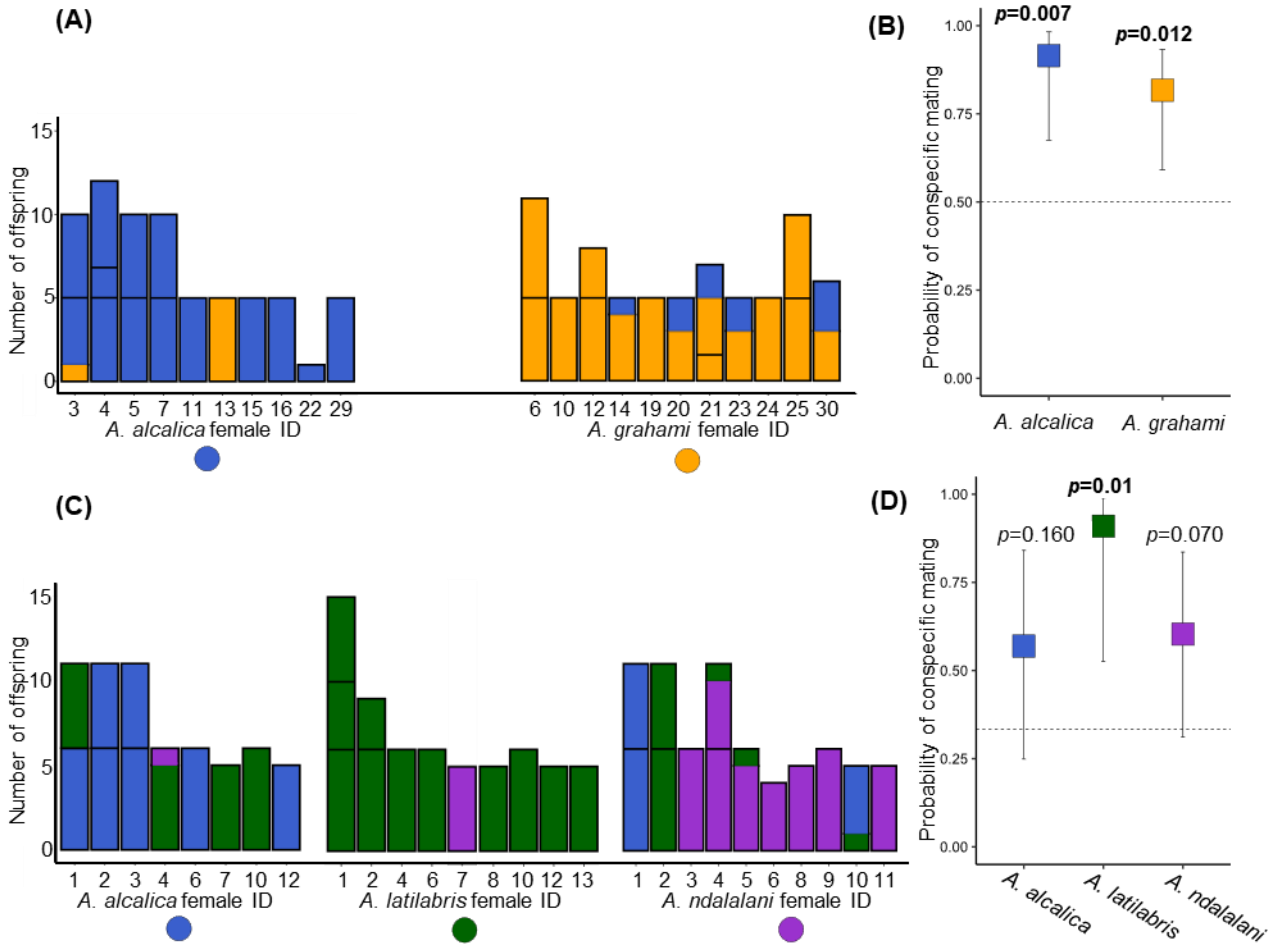


Figure 2.3. Assortative mate choice in *Alcolapia* recorded from offspring paternity data. Barplots show paternity of offspring for each spawning female in both allopatric (**A**) and sympatric (**C**) experiments. Colours denote the paternal species assigned from the microsatellite data: orange = *A. grahami*, blue = *A. alcalica*, green = *A. latilabris*, purple = *A. ndalalani*. Each column represents an individual female. Within columns, individual clutches are separated by black bars. (**B**, **D**) Boxplots showing the predicted probability of females mating with conspecific species from offspring paternity data, (**B**) allopatric experiment, (**D**) sympatric experiment calculated from the model outputs of the GLMs. Error bars denote 95% Wald confidence intervals. Dotted lines show the mean expected probability of mating with conspecifics given random mating. In both experiments, mating was assortative overall ($p < 0.001$). However, in the sympatric experiment, tests for assortative mating were not significant for *A. alcalica* (p=0.16) and *A. ndalalani* (p=0.07).

2.3.3 Assortative mate choice

In the allopatric experiment, a total of 140 offspring from 31 broods were successfully genotyped with paternity assigned to an individual male in all cases. The total number of breeding females for each species was 10 for *A. alcalica* and 11 for *A. grahami* (Figure 2.3A). The proportion of conspecific offspring was 88.6% (124/140) for all the offspring in the allopatric experiment. Both *A. alcalica* (62 of 68 offspring) and *A. grahami* females (62 of 72 offspring) showed a significant preference towards conspecifics (GLM *post-hoc* test*, A. alcalica*: *t.ratio* = 3.0, *df* = 18 *p* = 0.007; *A. grahami*: *t.ratio* = 2.8, *df* = 18, *p* = 0.012; Figure 2.3B; Table A2.9).

In the sympatric experiment, a total of 193 offspring from 36 broods were successfully genotyped with paternity assigned to at least the species level. The total number of breeding females for each species was 8 for *A. alcalica*, 9 for *A. latilabris* and 9 for *A. ndalalani* (Figure 2.3C). Of the successfully genotyped offspring, 89.6% (173/193) of offspring were assigned to individual males. The proportion of offspring assigned to individual males differed among species; 90.2% (44/59) in *A. alcalica*, 91.9% (57/62) in *A. latilabris* and 81.4% (57/70) in *A. ndalalani* (Figure 2.4B). The proportion of conspecific offspring was 71.0% (137/193) for all the offspring in the sympatric experiment: 63.9% (39/61) in *A. alcalica* females, 98.4% (57/62) in *A. latilabris* females and 58.6% (41/70) in *A. ndalalani* females (Figure 2.3C). Overall, females spawned with conspecific males significantly more than expected given random mating (GLM *post-hoc* test, *t.ratio* = 3.65, *df* = 23, *p* = 0.001; Table A2.9; Table A2.9). However, levels of assortativeness differed between species, with *A. latilabris* females having a significant preference towards conspecifics (GLM *post-hoc* test, *t.ratio* = 2.82, *df* = 23, *p* = 0.010), but *A. alcalica* (*t.ratio* = 1.45*, df* = 23*, p* = 0.160) and *A. ndalalani* females did not (*t.ratio* = 1.90, *df* = 23, *p* = 0.070; Figure 2.3D; Table A2.9)

Multiple mating was relatively common, with 61.31% (19/31) and 69% (25/36) of broods showing multiple paternity in the allopatric and sympatric experiments respectively (Figure 2.4). These represent the minimum level of multiple paternity as we only genotyped a maximum of six offspring per brood (brood sizes varied from 3 to 51), and we also could not distinguish between all males in the sympatric experiment. Individual mating success also varied among males. For instance, in the allopatric experiment, one *A. grahami* male sired at least one offspring with every

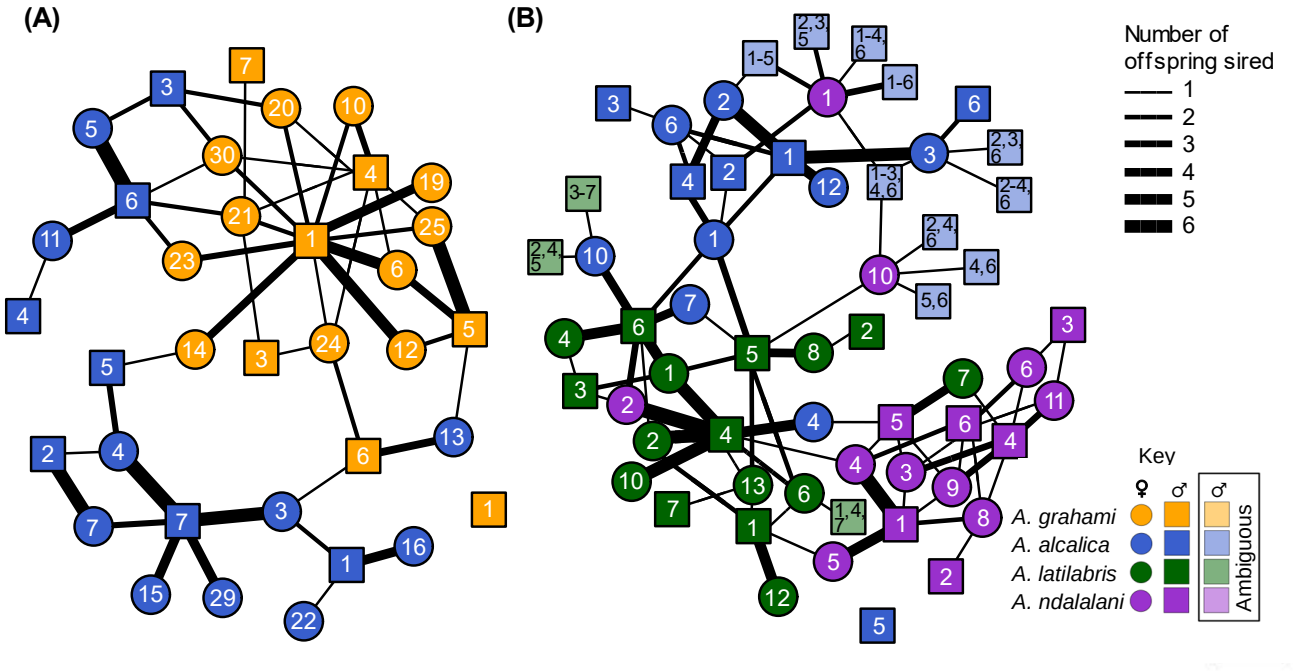


Figure 2.4. Network showing parent-offspring relationships in the allopatric experiment (**A**) and sympatric experiment (**B**) using offspring paternity data. Colours indicate species: blue = *A. alcalica*, orange = *A. grahami*, green = *A. latilabris*, purple = *A. ndalalani*. Circles = females and squares = males. Numbers correspond to unique male or female ID. Line thickness denotes the number of offspring sired between parents. Lighter colours denote offspring with ambiguous paternity that were not assigned to a single compatible male. Paternity could be assigned unambiguously in 94.0% of offspring. 88.6% (two-way choice) and 71.0% (three-way choice) of offspring in the allopatric and sympatric experiments resulted from conspecific matings. In the allopatric experiment, 3 females and one male produced no offspring; in the sympatric experiment, 4 females produced no offspring and a single male produced no (unambiguously assigned) offspring.

breeding *A. grahami* female, but none with any *A. alcalica* females (Figure 2.4A). By contrast, a single *A. grahami* male and *A. alcalica* male were not assigned parentage to any offspring in the allopatric and sympatric experiments, respectively (Figure 2.4A, B).

2.4. Discussion

2.4.1 Main findings

We demonstrate that *Alcolapia* species show some evidence of assortative mating in comparisons both between sympatric and allopatric species in a semi-natural aquarium set-up. Hybrid offspring were generated between all species pairs that were tested, indicating that in our experimental setup, reproductive isolation is incomplete. Together with results from previous studies showing differences in trophic morphology between sympatric species (Ford *et al.,* 2016), our findings lend support to the theory that speciation in sympatry is more likely when there is a combination of assortative mating and ecological divergence (Dieckmann and Doebeli, 1999). These findings correspond with other studies of recent and small cichlid adaptive radiations, such as Lake Ejagham in Cameroon and the crater lakes of Nicaragua (Martin, 2013; Machado-Schiaffino *et al.,* 2017).

2.4.2 The evolution of assortative mating in allopatry and sympatry

The forces driving speciation may vary depending on the geographic context and amount of gene flow between populations. In general, it is expected that prezygotic barriers will be stronger among sympatric compared to allopatric taxa as there is selection to avoid heterospecific matings in sympatry, but not in allopatry (Butlin, 1987; Coyne and Orr, 1997). Contrary to these expectations, our results could not detect a difference in the degree of assortative mating between allopatric species and sympatric species.

While reinforcement may be a key driver of premating isolation in some taxa (Yukilevich, 2012), premating barriers are also predicted to increase with genetic distance (Zouros, 1973; Coyne and Orr, 1989). Allopatric *Alcolapia grahami* females showed strong premating isolation, but they are also more genetically distinct compared to the sympatric Natron species (Figure 2.1; Ford *et al.,* 2015). On the other hand, phenotypic rather than genetic distance is often a better predictor of assortative mate choice (McPeek and Wellborn, 1998). For example, in *Pseudocrenilabrus* spp. of Lake Mweru and Bangweulu, less closely related but more phenotypically similar species have stronger premating isolation than more distantly related but less

phenotypically similar species (Stelkens *et al.,* 2009). *A. grahami* are more phenotypically similar to *A. alcalica* in terms of trophic morphology (head and mouth shape) compared to the other sympatric Natron species (Ford *et al.,* 2016), but they do differ in male colouration, which is typically a primary cue used by females for mate selection (Seehausen *et al.,* 1997; Selz *et al.,* 2014). A meta-analysis of cichlid premating isolation studies found only slight differences in the levels of assortative mating between sympatric and allopatric species (Rometsch *et al.,* 2020). However, they also found greater variance in premating isolation among allopatric species and lower levels when comparing allopatric and sympatric populations, suggesting an important role in the duration of the population split. Nevertheless, strong assortative mate choice can evolve among cichlids over relatively short periods of allopatry (Knight and Turner, 2004; Genner *et al.,* 2007), especially if there are divergent ecological pressures resulting from habitat differences (Tyers *et al.,* 2014).

Fossil and geological data date the Magadi-Natron split to approximately 8 Ka (Dommain *et al.,* 2022), and the main sections of the lakes are currently situated approximately 25 km apart (Williamson *et al.,* 1993; Figure 2.1A). Using nuclear genetic data, the *A. grahami* (Magadi) – Natron *Alcolapia* species split has been estimated at 0.007-1.55 Ma (95% HPD) (Ford *et al.,* 2019), which suggests that *A. grahami* populations in Magadi have been separated for at least 7 ka and may have diverged before the lakes split (Dommain *et al.,* 2022; Williamson *et al.,* 1993). Our results, therefore, indicate that strong mate discrimination can either evolve or persist in allopatry without recent (< 7 ka) reinforcement.

The weak species-assortative mating observed in sympatric *A. alcalica* and *A. ndalalani* in this experiment differs from other cichlid studies which generally find strong premating isolation among sympatric species (Knight and Turner, 1998; Plenderleith *et al.,* 2005; Machado-Schiaffino *et al.,* 2017). However, levels of assortative mating can be lower and more variable among more recently separated sympatric cichlid species or populations (Jordan *et al.,* 2003; Selz *et al.,* 2016; Nyalungu and Couldridge, 2020). For instance, Selz *et al.* (2016) found varying levels of assortative mating between *Pundamilia nyereri* populations in two-way choice experiments, but mate choice was strongly assortative when females were provided with a choice of closely related *P. igneopinnis* (96-100%) (Selz *et al.,* 2016). Premating

isolation in sympatric *Alcolapia* may therefore be more comparable to that observed between populations or sub-species in other cichlid systems. Incomplete assortative mating in *Alcolapia* may be unsurprising given that the radiation is extremely recent (Ford *et al.,* 2015) and that the degree of genetic differentiation is among the lowest measured across all cichlid radiations (Svardal *et al.,* 2021). Moreover, asymmetry in mating preferences or species discrimination may be common among recently diverging cichlids (Nevado *et al.,* 2011; Malinsky *et al.,* 2015; Van Steenberge *et al.,* 2021) and hence it may be unsurprising that some species exhibit weaker assortative mate choice.

2.4.3 Factors influencing the strength of assortative mate preferences

The levels of hybridisation among *Alcolapia* species reported in our experiments (Figure 2.3) are high enough that species would collapse within a few generations given negligible postzygotic effects (Irwin and Schluter, 2021). Yet both genetic (Figure 2.1) and morphological data indicate that hybrids are comparatively rare in most wild populations (Ford *et al.,* 2015; Ford *et al.,* 2016), suggesting that premating isolation is either stronger in the wild or is accompanied by strong selection against hybrids during early life stages.

Assortative mate choice may be influenced by extrinsic factors such as spatial, temporal and environmental components, some of which may not be present in an aquarium setup. For example, spatially-mediated size-assortative mating has been observed in the cichlid *Eretmodus cyanostictus*, where larger males dominate high-quality habitats while smaller, subdominant males occur more frequently in low-quality environments (Taborsky *et al.,* 2014). In the wild, *Alcolapia* may exhibit some spatial separation which could affect encounter rates and influence levels of assortative mating. For instance, Seegers *et al.* (2001) recorded a higher abundance of *A. latilabris* in the upper courses of streams. During field collections, we observed that *A. latilabris* and *A. ndalalani* mainly occurred in upstream sections and in more rocky habitats, whereas downstream sections with fine-grained substrates were often dominated by *A. alcalica*. On the other hand, observations of breeding leks in Lake Natron were found to be comprised of multiple different species (Seegers and Tichy, 1999). Furthermore, the occurrence of multiple allopatric sites containing only *A.*

alcalica indicates that its area of sympatry is less extensive compared to *A. latilabris* and *A. ndalalani*, which have a significant overlap in their distribution (Figure 2.1A).

A multitude of ecological and environmental factors may influence premating isolation beyond the primary cues used for mate choice. While previous studies on the sensory cues used by cichlids in assortative mate choice have usually found visual cues such as male colouration to be the primary premating cues (Seehausen *et al.,* 1997; Selz *et al.,* 2014), single cues alone seldom control all premating isolation (Plenderleith *et al.,* 2005; Blais *et al.,* 2009; Rometsch *et al.,* 2020). Instead, sensory cues may be multimodal, with each cue contributing to premating isolation to different degrees (Houck and Verrell, 1993; Rafferty and Boughman, 2006; Mérot *et al.,* 2015; Keller-Costa, 2015). Due to the lack of shade from terrestrial vegetation, the shallow depths and the high clarity of the water column, the light environment of Natron and Magadi is extremely bright (Johannsson *et al.,* 2014). This bright visual environment may not be fully replicated in the aquarium setup and subsequently, visual cues involved in mate choice may become less salient (Maan *et al.,* 2010; Wright *et al.,* 2018). In line with this hypothesis, the breakdown of reproductive barriers via the loss of visual cues may already have occurred in *Alcolapia*, with a high turbidity site on the eastern shore of Natron supporting a potential hybrid population with intermediate morphology (Ford *et al.,* 2015).

Olfactory cues have also been shown to be a component of mate choice in some cichlid species (Plenderleith *et al.,* 2005, Blais *et al.,* 2009). Partitioning of diet among *Alcolapia* species could potentially promote the differentiation in olfactory cues used in mate choice (Kavembe *et al.,* 2016; Ford *et al.,* 2016). As species were fed an identical diet in these experiments, any diet-related odour discrimination would be eliminated. The combination of ecologically mediated premating cues may have an additive effect, where assortative mating is relatively weak with only primary mate choice cues but strong in the presence of additional factors (Tinghitella *et al.,* 2020). Future studies may investigate the relative contribution that different cues (visual, olfactory etc.) play in *Alcolapia* reproductive isolation with aquarium experiments (e.g. Knight and Turner, 1999; Selz *et al.,* 2014). Other factors relating to population density and spatial structuring that may also influence mate choice should be considered.

CHAPTER 3: DIET AND MICROBIOME NICHE PARTITIONING AS A DRIVER OF DIVERSIFICATION IN THE *ALCOLAPIA* ADAPTIVE RADIATION

ABSTRACT

Ecological niche partitioning is a key feature of adaptive radiations and typically involves the diversification of diets and habitats. Both diet and habitat significantly influence gut microbial communities, meaning that microbiome diversification may coincide with ecological partitioning. Soda lake cichlids (*Oreochromis* (*Alcolapia*) spp.) are notable for their adaptation to extreme conditions and rapid diversification in Lake Natron, Tanzania. The three Natron *Alcolapia* species display extensive ecomorphological and dietary niche diversification despite inhabiting a harsh environment lacking in trophic diversity. To explain the co-existence of species with similar trophic niches, we conducted metabarcoding of gut contents to investigate the diet and gut microbiome of *Alcolapia*. Using amplicon sequencing of algae, invertebrates, and bacteria, we discovered cryptic diversity in diets and identified microbial taxa present in the gut. Compared to other cichlid radiations, we find that the gut microbiome of *Alcolapia* lacks some core cichlid bacterial taxa but has more genera associated with halophilic and alkaliphilic growth. The diet of all three species is dominated by cyanobacteria, the most abundant of which could only be identified to the level of class or phylum, indicating high levels of uncatalogued algal diversity in this lake. We found significant differences in the diet and microbiome between the three species, although most of the variation was explained by site differences. Our findings indicate that the divergence in *Alcolapia* morphology may be linked to utilisation of the same resources using alternate foraging techniques, with the exploitation of different resource types as a secondary factor. Differences in host species gut bacterial composition were more prominent than diet-associated differences, suggesting that microbial communities may be a better indicator of host species identity than diet among closely related species occupying the same trophic level.

3.1 Introduction

Ecological processes are central drivers of adaptive radiation and occur when barriers to gene flow arise as a consequence of ecologically divergent selection (Schluter, 2001; Rundle and Nosil, 2005). When populations are in contact, the partitioning of resources can reduce competition between diverging lineages and enable species to coexist (Schoener, 1974; Turelli *et al.,* 2001). In many vertebrate lineages, a principal mechanism of niche segregation is through trophic partitioning. This can promote reproductive isolation through the divergence of ecological phenotypes or if populations exploit adjacent habitats (Streelman and Danley, 2003; Ackerly *et al.,* 2006). Documenting the diet and habitat differences between diverging taxa, as well as their associated ecological phenotypes, is therefore essential for studying ecological speciation.

Diet and habitat are also correlated to the structure of gut microbial communities, making the gut microbiome a potential signature of ecological partitioning (Ley *et al.*, 2008; Miyake *et al.*, 2015; Loo *et al.*, 2019). There is also increasing evidence that host-associated microbial communities can have a significant effect on hosts' fitness (Rosshart *et al.*, 2017; Gould *et al.*, 2018) and can provide a source of phenotypic novelty for selective pressures to operate (Henry *et al.*, 2021). This symbiotic relationship between the host and the microbiome therefore has implications for speciation (Brucker and Bordenstein, 2012; Shropshire and Bordenstein, 2016). Characterising the host microbiome and investigating the interactions between the host and environment can provide additional insights into processes underlying diversification among radiating lineages.

Cichlid fishes from the East African great lakes are famous examples of adaptive radiation and rapid speciation (Turner *et al.*, 2001; Salzburger and Meyer, 2004). Ecological partitioning is thought to have played a major role in the diversification of these systems (Kocher, 2004, Wagner *et al.,* 2012), but there are also examples where co-occurring species with divergent morphologies show considerable dietary overlap (Genner *et al.*, 1999a; Genner *et al.,* 1999b; Martin and Genner, 2009, Hata *et al.,* 2014). Cichlid gut microbiome diversification has been studied in Lake Tanganyika, an ancient rift lake comprising multiple radiations, as well as in younger (typically < 1 Ma)

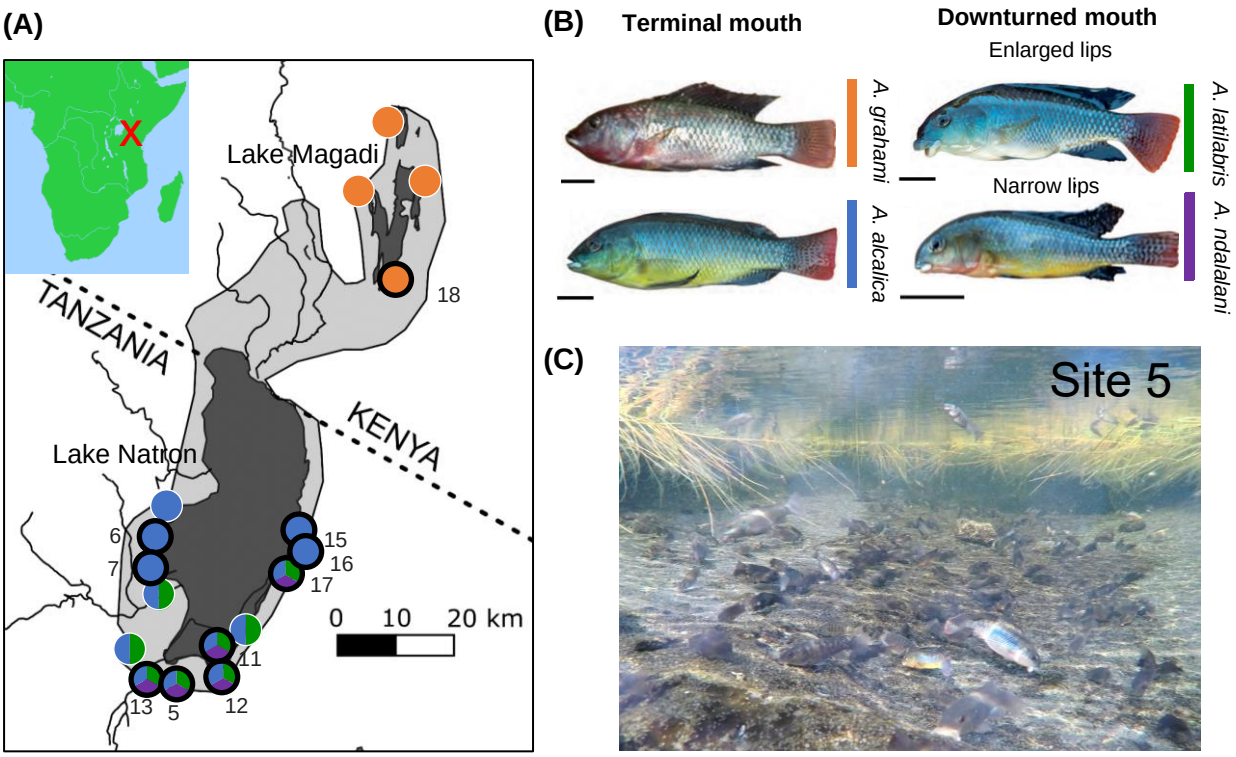


Figure 3.1. Distribution, morphology and ecology of *Alcolapia*. (**A**) Map showing the sites selected for *Alcolapia* gut metabarcoding. The light grey area shows the extent of Palaeolake Orolonga (ancient freshwater lake containing *Alcolapia* common ancestor), coloured circles correspond to species in (**B**). Samples and sites used in this study are labelled and outlined with black circles. Site 11 is a structured site and contains 3 sub-sites (11A, 11B and 11C). Sites 5, 11, 12, 13 and 17 contain all 3 sympatric species. (**B**) Lake Magadi contains *Alcolapia grahami* while Lake Natron contains *A. alcalica*, *A. ndalalani* and *A. latilabris.* Extensive differences in mouth shape are observed between species. Terminal mouth = *A. alcalica* and *A. grahami*, downturned mouth = *A. latilabris* and *A. ndalalani. A. latilabris* also has a wider mouth and enlarged lips. Scale bar = 1cm. (**C**) A photograph of spring site 5 containing sympatric *Alcolapia* species. The *Alcolapia* habitat is typically very shallow (<1m depth) and species are highly locally abundant. In the foreground are colourful male *A. ndalalani* (left) and *A. latilabris* (right). Partially submerged sedges (*Cyperus*) occur along the littoral margins of this site. Schematic map and fish photograph are taken and adapted from Ford *et al.,* 2015. Site 5 photo credit: Julia Day.

and less diverse (< 15 species) systems such as crater lakes (Franchini *et al.*, 2014; Baldo *et al.*, 2017, 2019; Härer *et al.*, 2020; Riera and Baldo, 2020; Mejía *et al.*, 2022). Across these systems, gut microbiota composition was found to vary between lakes as well as between cichlid species occupying different trophic levels, such as carnivores and herbivores. However, gut microbiota differences between closely

related sympatric species were either absent or minimal (Franchini *et al.*, 2014; Baldo *et al.*, 2019; Härer *et al.*, 2020). Despite being separated by large phylogenetic and geographic distances, cichlids from Africa and Central America were found to share a core set of microbial taxa (Baldo *et al.*, 2019).

Alcolapia soda lake cichlids are an ideal system in which to uncover the role that dietary and gut microbiome niche differentiation plays in speciation. As a small (4 described species) and recent (< 10 kya) adaptive radiation restricted to two extreme alkaline, soda lakes in East Africa (Lake Natron and Magadi), *Alcolapia* cichlids are still at the early stages of speciation (Seegers and Tichy, 1999; Ford *et al.*, 2015). The three species present in Lake Natron coexist sympatrically at some of the same spring sites and show incomplete reproductive isolation (Ford *et al.*, 2015; Chapter 2; Figure 3.1). For such a young species flock, the species show extensive divergence in trophic morphology, with a terminal mouth species (*Alcolapia alcalica*) and two downturned mouth forms, with enlarged lips (*Alcolapia latilabris*) and narrow lips (*Alcolapia ndalalani*) (Seegers and Tichy, 1999; Ford *et al.*, 2016; Figure 3.1B).

This morphological diversity is present despite *Alcolapia* inhabiting a shallow (< 1 m) and depauperate environment that precludes diversification at different depths and trophic levels, which is typical of other cichlid radiations (Seehausen *et al.*, 2008; Wagner *et al.,* 2012; Figure 3.1C). Stable isotope analysis (SIA) and visual examination of gut contents found that the species are largely algivorous and occupy a similar herbivorous trophic niche (Coe, 1966; Ford *et al.*, 2016), but small differences between species indicated some degree of very fine-scale niche differentiation (Ford *et al.*, 2016). In particular, the terminal mouth species *A. alcalica* exhibited a shifted carbon isotope profile, longer guts and a higher proportion of cellulose in its stomach contents, indicative of feeding on some vascular plants (Ford *et al.*, 2016). However, differences between benthic species (*A. ndalalani* and *A. latilabris*) were minor: both species had very similar stable isotope profiles and only small differences in stomach contents, with *A. latilabris* having a higher proportion of grit (Ford *et al.*, 2016). It was hypothesised that niche differences between these two species were likely behavioural – *A. latilabris* exhibiting a grazing foraging strategy, while *A. ndalalani* a browsing strategy (J.J. Day, personal communication, 21st November 2022). However, as traditional analysis of *Alcolapia* stomach contents was only able to

resolve dietary taxa to very broad categories (e.g. “algae”) there may be unidentified cryptic diversity regarding dietary preferences between cichlid species.

In addition to studying the diet, the hostile conditions of Natron and Magadi spring sites also make for an novel case study of microbiome diversification in an extreme environment. *Alcolapia* have long been of research interest due to their many unique physiological adaptations to survive extremes of pH, temperature, salinity, UV light and variable oxygen levels (Trewavas, 1983; Narahara *et al.*, 1996; Walsh *et al.*, 2001; Wood *et al.*, 2012, 2016, White *et al.,* 2020). Physiological adaptations of the gut are one such innovation, where a branching of the oesophagus prevents alkaline water entering the stomach and interfering with gastric acids (Bergman *et al.*, 2003). These extreme conditions may also have implications on the structure of gut microbiota, which is known to play a role in hosts’ survival by regulating metabolic pathways (Debelius *et al.*, 2016; Wang *et al.*, 2018). The composition of gut microbiota is also strongly influenced by microbes present in the external environment, especially among fishes and other aquatic organisms (Ghanbari *et al.,* 2015; Smith *et al.*, 2015; Kim *et al.*, 2021). However, the gut microbiota of fishes found in extreme environments has rarely been studied, but a few studies exist (Bhute *et al.*, 2020; Sepulveda and Moeller, 2020).

Molecular methods present the most promising approach to characterise both the diet and gut microbial composition of *Alcolapia*. Stomach content analysis and SIA are limited in their ability to identify the diet of algivorous species exhibiting fine-scale differences (Ford *et al.*, 2016). Metabarcoding has previously been applied to fish stomach and gut contents to explore the diversity of prey taxa when visual examination is problematic (Leray *et al.,* 2015; Jakubavičiute *et al.*, 2017; Casey *et al.*, 2019). The same approach has also successfully been used to identify cryptic diet variation between algivorous fishes, such as in Lake Tanganyikan cichlids that defend algae farms (Hata *et al.*, 2014) and in Hawaiian surgeonfish (Nalley *et al.,* 2021).

In this study, we characterised both the diet and gut microbiota of *Alcolapia*. Our main aims were to:

1. Characterise the composition of the *Alcolapia* gut microbiome and diet to a high taxonomic resolution.

2. Assess the extent that *Alcolapia* gut microbiota differ from other cichlid radiations, testing the hypothesis that extreme aquatic conditions will lead to a core microbiome divergent to that of other cichlids and enriched for extremophile-associated microbial taxa.

3. Test if sympatric *Alcolapia* species exhibit cryptic differences in diet and gut microbiota indicative of resource partitioning.

3.2. Methods

3.2.1 Sample collection

Specimens of all three *Alcolapia* species were previously collected by collaborators from sites across Lake Natron (*A. alcalica A. ndalalani*, *A. latilabris*), Tanzania, (Figure 3.1). These specimens were collected in January 2012, August 2012 (permit number: 012-25-NA-2011-182) and July 2017 (permit number: 2017-259- NA-2011-182). Species were identified in the field using morphological features (Seegers and Tichy. 1999). Fish were euthanised using clove oil before being preserved in 95% ethanol. Gut dissections and DNA extractions were carried out in June 2021.

To explore differences between species and sites, 220 specimens were selected from multiple sites across Lake Natron. This included sites containing all three sympatric species (sites 5, 11A, 11B, 11C, 12, 13) and sites containing only *A. alcalica* (sites 6, 7, 15A, 15B, 16; Figure 3.1; Table 3.1). Site 11 encompassed a stream at the south of Lake Natron and consisted of three subsites with varying conditions (11A = lagoon on lake edge, 11B = fast flowing stream with small rapids, 11C = spring source). 15B was a small pond separate to site 15A and contained *A. alcalica* with upturned mouth morphology (Ford, 2015).

3.2.2 Specimen dissection and DNA extraction

All instruments used for dissections were disinfected between samples by rinsing in a 1:200 solution of Distel Trigene disinfectant (Tristel Solutions Ltd., UK) for five minutes, before rinsing with $_{dd}H_20$. Specimens were first rinsed in 70% ethanol. The entire gastrointestinal tract from below the trifurcation point of the stomach to the anus was

Table 3.1. Number of *Alcolapia* specimens used for metabarcoding of gut extracts. For each species, the numbers used from each site at both collection years are provided. Grey shaded areas denote species that have not been recorded at a given site during previous collections.

Year	Site	Habitat	No. of endemic species	*A. alcalica*	*A. latilabris*	*A. ndalalani*
2012	5	Spring	3	1	10	11
	11A	Lagoon		9	-	10
	11B	Rapids		7	6	6
	11C	Spring		-	8	8
	12	Spring		6	8	9
	13	Spring		5	5	1
	6	Spring	1	10		
	7	Spring (structured)		8		
	16	Spring		11		
2017	5	Spring	3	10	5	2
	12	Spring		9	-	-
	17	Spring		5	15	8
	7	Spring	1	6		
	15A	Spring		11		
	15B	Pool		10		

used for DNA extractions, which therefore included both the gut contents (digesta) and gut lining. This enabled sampling of dietary sources as well as microbes of autochthonous (native to the gut lining) and allochthonous (derived from the environment) origin (Ringø *et al.*, 2016). Gut fullness of specimens was estimated by unravelling the gastrointestinal tract and measuring the sections containing digesta as a proportion of the total length.

Table 3.2. Amplicon primers selected to target dietary (plant/algae - 23S plastid rRNA, Macroinvertebrates – COI B5) and microbiome (16S rRNA V3-V4) taxa in *Alcolapia* guts. The PCR conditions of the first stage PCR are provided.

Target region	Primer name	Direction	Sequence (5' - 3')	Target taxa	Amplicon length (bp)	First stage PCR conditions	Reference
Region V3–V4 16S rRNA	S-D-Bact-0341-b-S-17	Forward	CCT ACG GGN GGC WGC AG	Bacteria (microbiome)	~464	95 ˚C for 15 min, 25 cycles of 94 ˚C for 30 sec, 53 ˚C for 90 sec, and 72 ˚C for 1 min, and a final extension at 72 ˚C for 10 min	Klindworth *et al.,* 2013
	S-D-Bact-0785-a-A-21	Reverse	GAC TAC HVG GGT ATC TAA TCC				
Region V5 23S rRNA	p23SrV_f1	Forward	GGA CAG AAA GAC CCT ATG AA	Plants and algae (eukaryotic and cyanobacteria)	~410	95 ˚C for 15 min, 25 cycles of 94 ˚C for 30sec, 63 ˚C for 90 sec, and 72 ˚C for 1 min, and a final extension at 72 ˚C for 10 min	Sherwood and Presting, 2007
	p23SrV_r1	Reverse	TCA GCC TGT TAT CCC TAG AG				
COI B5	B	Forward	CCI GAY ATR GCI TTY CCI CG	Macroinvertebrate prey	310	95˚C for 15min, 30 cycles of 94˚C for 30sec, 46˚C for 90sec, and 72˚C for 1min, and a final extension at 72˚C for 10min	Hajibabaei *et al.,* 2019
	ArR5	Reverse	GTR ATI GCI CCI GCI ARI ACI GG				

Prior to DNA extraction, the gut tissue lining of each specimen was disrupted using a pestle and mortar following freezing using liquid nitrogen. Up to 200 mg of this gut sample was used for extraction using a DNEasy® PowerSoil® Pro extraction kit (Qiagen), the precursor of which has been used successfully in a fish gut metabarcoding study (Nalley *et al.*, 2022). To aid cell lysis, an additional incubation step (10 min at 65 °C) was added and then samples were homogenised using a TissueLyser II (Qiagen, Venlo, Netherlands). DNA was eluted into 50 µl. Following quantification using an Invitrogen™ Qubit™ 3 Fluorometer (Thermo Fisher Scientific), all DNA extracts were normalised to a concentration of 10 ng/µl. Two extraction controls were also included to monitor sources of contamination occurring from the extraction stage onwards.

3.2.3 Amplicon primer selection

To explore diversity in plant and algae dietary species we used the p23SrV_f1 and p23SrV_r1 primers which target the 23S ribosomal RNA (rRNA) Universal Plastid Amplicon (UPA) (Sherwood and Presting, 2007; Table 3.2). *Alcolapia* are predominantly herbivorous, and therefore may be feeding on a combination of cyanobacteria, green algae, diatoms, and plants (Coe 1966; Ford *et al.*, 2016). As this encompasses both prokaryotic (cyanobacteria) and eukaryotic species, the 23S UPA primers enable approximate estimates of the relative contribution of these different algal taxa to diet. The cytochrome oxidase I (COI) B and ArR5 primers were used to target macroinvertebrate prey taxa and other metazoans (Hajibabaei *et al.*, 2019) (Table 3.2). To explore the diversity of gut and dietary microbes, we used the bacterial/prokaryotic primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21 (Klindworth *et al.*, 2013; Table 3.2). This primer pair amplifies the V3–V4 region of 16S ribosomal rRNA and targets bacteria, including both autochthonous and allochthonous microbes. All primers were tailed with Illumina adapters: forward adaptor 5'-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-3'; reverse adaptor 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3'.

3.2.4 Library preparation

Annealing temperatures, numbers of PCR cycles, amounts of template DNA and primer concentrations were first optimized for each primer pair (Table 3.2). PCRs were carried out in 20 µl reaction volumes using the Qiagen Multiplex PCR kit (Qiagen) and containing 10 µl of Multiplex 2x PCR Master Mix. For 16S, the reaction contained 0.375 µM of each primer and 10 ng of DNA extract. For 23S and COI, the reactions contained 0.5 µM of each primer and 20 ng of DNA. Each 96-well plate contained the two extraction controls and six randomly placed blank controls to monitor contamination, with a total of three plates of samples per primer pair. PCRs were also repeated for 24 samples (~ 11%) for troubleshooting purposes and to assess the repeatability of the results.

To reduce the number of poorly amplified samples and amplification bias effects, two separate PCRs were carried out for each sample for each primer pair. The PCR products from each of these separate reactions were then combined for subsequent

steps. PCR products were purified using AMPure XP beads (Beckman Coulter, Netherlands).

A final PCR using Illumina Fi5 and Ri7 primers (Integrated DNA Technologies, Iowa, USA) was used to add unique identifier sequences to allow sample pooling prior to sequencing, with the following cycling parameters: initial 95 °C for 15 min, followed by 8 cycles at 98 °C for 10 sec, 65 °C for 30 sec, 72 °C for 30 sec and a final annealing step of 72 °C for 5 min. Following this PCR, all sample concentrations were quantified using a FLUOstar OPTIMA (Promega, Wisconsin, USA), pooled in equimolar groups of 12, and purified using AMPure XP beads. These pools were quantified by qPCR (KAPA Library Quantification Kit, Kapa Biosystems, USA). DNA from these pools were then combined in equimolar amounts to a concentration of 4 nM, with a single library for all COI samples and another library containing all 23S and 16S samples. These two libraries were sequenced using separate 2 x 250bp paired end runs using an Illumina MiSeq v2 (Ilumina, California, USA) at the Centre for Genomic Research, University of Liverpool, UK.

3.2.5 Read processing and taxonomic assignment

Illumina adapter sequences were trimmed from the raw fastq files using *Cutadapt* v.1.2.1 (Martin, 2011) with the option "-O 3". Sickle v.1.200 (Joshi and Fass, 2011) was used to trim reads with a minimum window quality score of 20 and to remove reads shorter than 15bp.

The *DADA2* v.1.24.0 *R* package was used to process unpaired reads and generate amplicon sequence variants (ASVs) (Callahan *et al.*, 2016; R Core Team, 2022). ASVs were used in place of operational taxonomic units as they capture taxonomic variation at a finer resolution and have greater reproducibility (Callahan *et al.*, 2017). In *DADA2,* the command filterandtrim was used to remove sequences containing ambiguous bases, filtered with options "truncQ=2, maxEE=c(2, 2)", and discard reads with phiX sequence. Primers were trimmed from reads using *cutadapt* v.3.4 and reads below 60bp were discarded (Martin, 2011). After identical reads were dereplicated, ASVs were inferred using *DADA2*'s error model with the inference algorithm run on single samples. Forward and reverse paired end reads were merged using the default minimum overlap of 12bp. After creating an ASV matrix, chimeric sequences were

removed using the “removeBimeraDenovo” command with ‘method = “consensus”’. ASVs were filtered for minimum and maximum sizes for each dataset (16S: >350 and < 430bp, 23S: >350bp and < 407bp, COI: >300bp and <330bp).

For 16S rRNA microbial amplicons, taxonomy was assigned using the *IDTAXA* algorithm from the *DECIPHER* v 2.24.0 *R* package (Murali *et al.*, 2018) with default parameters and using the prokaryotic SILVA_SSU_r138_2019 training set (https://zenodo.org/record/4587955#.YtGRqexBz0o). Further resolving of microbial taxa to exact species was carried out using exact matching using *DADA2*’s “addSpecies” command and the silva_species_assignment_v138.1 training set (Callahan *et al.*, 2016). For COI, sequences were searched against a curated *BLAST+* database containing all sequences assigning to “cytochrome” or “mitochondria” in GenBank (Camacho *et al.*, 2009). Results were filtered for > 90% identity, e value < 0.001 and alignment length of at least 100 bp. *BLASTn* results were input into *MEGAN* v6.22.2, where taxonomy was assigned using the “megan-nucl-Jan2021.db” mapping file and the LCA algorithm (Huson *et al.*, 2016). For 23S rRNA sequences, taxonomy was assigned using the *DADA2* “assignTaxonomy” function with a bootstrap cut-off of 80% (default = 50%) for assigning taxonomy and the µ-green 23S rRNA plastid database “µgreen-db_r1.1” (Djemiel *et al.*, 2020). This curated database contains a wide assemblage of algal species, but few non-algal plant taxa. Non-algal plant ASVs were resolved further (up to genus level) using *BLASTn* on GenBank (Camacho *et al.*, 2009), followed by *MEGAN* (Huson *et al.,* 2016).

3.2.6 Filtering of contaminants, non-target taxa and normalisation

For each amplicon dataset, the ASV table, taxonomy and sample metadata was combined to make a *phyloseq* v.1.40.0 object in *R* (McMurdie *et al.*, 2013). Contaminants were filtered using the frequency and prevalence method in the *decontam* package v. 1.16.0 (Davis *et al.*, 2018). The ASVs were aligned using *DECIPHER* v2.0.2 (Wright, 2016) and input to *FastTree* v.2.1.1 (Price *et al.,* 2010) to generate a phylogenetic tree. For all datasets, any ASVs that were not assigned to a kingdom were removed. For the 16S rRNA dataset, reads unassigned at phylum level or that were assigned to “Chloroplast”, “Cyanobacteria”, “Mitochondria” or “Other” were also discarded. For the COI dataset, only ASVs that were assigned to the kingdom “Metazoa” were kept. All subsequent analyses were conducted in *R*.

For alpha and beta diversity analysis, samples were normalised to the same sequencing depth using the *R* package *Scaling with Ranked Subsampling* (*SRS*) v.0.2.3, which minimises artefacts caused by variable library sizes (Beule and Karlovsky, 2020). For *SRS* normalisation, a C_{min} (total read count normalisation value) threshold of 1061 reads was chosen for the 23S amplicon, which only led to 8 samples being dropped but 92.1% of global species richness was retained (Figure A3.1). After removal of cyanobacterial reads, a C_{min} of 1145 was used for the 16S rRNA amplicon, which dropped 25 samples and retained 79.5 % of global species richness (Figure A3.2). For the COI amplicon dataset, a C_{min} 1378 was selected, which retained 88.0% of global species richness and dropped 13 samples (Figure A3.1). Of these dropped samples, 10 were 'upturned' *A. alcalica* that had many unassigned reads, which was potentially a result of amplification of reads from non-target taxa.

3.2.7 Characterisation of core bacterial taxa and extremophile taxa

Samples that had over 10,000 reads after removal of all cyanobacterial reads (n=56) were used to explore core *Alcolapia* gut microbial communities. To qualitatively compare *Alcolapia* core bacterial taxa with those of multiple other cichlid radiations, we used datasets from Baldo *et al*. (2019) (n=266). As per the authors' description (Baldo *et al.,* 2019), the OTU table was filtered by removing OTUs with less than 3 counts per specimen and less than 50 counts in total. To ensure taxonomic consistency between the datasets, the taxonomy of *Alcolapia* ASVs was re-assigned using the Greengenes gg_13_8 database used by Baldo *et al*. (2019) using *DADA2*'s assignTaxonomy algorithm with an 80% minimum bootstrap support. The Baldo *et al.* (2019) dataset was then split into their three respective systems to compare to *Alcolapia*: 1. Lake Tanganyika, 2. Barombi Mbo and 3. Nicaraguan crater lakes, selecting only the smaller lakes (< 20 km^2) for comparison (Apoyo, Masaya, Xiloá, Apoyeque, Asosca León) (Barluenga and Meyer, 2010). For each of these lake systems, core microbial taxa were identified at each taxonomic level by identifying only those taxa that had reads present in at least 90% of specimens. The proportion of samples harbouring these core taxa were then compared between systems.

To compare the prevalence of extremophile microbial taxa in *Alcolapia* guts to other cichlid species from the Baldo *et al.* (2019) dataset, the top 20 most abundant genera were selected from each dataset. For each genus, a list of described species was

retrieved using the LPSN (List of Prokaryotic names with Standing in Nomenclature) web interface (Parte *et al.,* 2020). Different growth characteristics were searched for in the literature with a species designated as alkaliphilic if it had optimal growth above pH 8 and/or a growth range above pH 9 and halophilic if the growth range required at least 2% NaCl (w/v) (Corral *et al.,* 2019). If species' growth conditions were not reported, they were listed as non-halophilic and non-alkaliphilic if derived from gut samples and halophilic if derived from marine environments or grown on marine agar. The genus was then labelled as having a growth trait common to either: 1. all its described species, 2. to some, but not all described species or 3. to none of its described species.

3.2.8 Diversity analyses

To examine differences in the microbiome (16S rRNA) and dietary phototrophic (23S rRNA) composition between sympatric species and sites, distance matrices were created using the non-phylogenetic Bray-Curtis dissimilarity and unweighted phylogenetic UniFrac (using *vegan*) and Generalised UniFrac (α = 0.5) metrics (GUniFrac v1.6) (Chen *et al.*, 2012; Oksanen *et al.*, 2022). The Bray-Curtis metric considers differences in relative abundance between ASVs, while the unweighted UniFrac incorporates phylogenetic distance but not abundance and is therefore more capable of detecting differences in rare taxa. Meanwhile, Generalised UniFrac is an extension to the unweighted UniFrac and weighted UniFrac (which accounts for abundance) phylogenetic metrics but has greater power in detecting differences between moderately abundant lineages (Chen *et al.*, 2012).

To observe relationships between sympatric species and sites, non-metric multidimensional scaling (NMDS) plots were generated using each distance metric using, *vegan* v.2.6-2 (Oksanen *et al.*, 2022). To test for differences in diet and microbiome composition, Permutational Multivariate Analysis of Variance Distance Matrices (PERMANOVA) was carried out using the "adonis" function in *vegan* (Warton *et al.*, 2012). As PERMANOVA has been found to yield unreliable results when performed on unequal samples sizes and heteroscedastic data, *p*-values for tests of these differences were also calculated using a distance-based Welch MANOVA test (W^*_d) (Hamidi *et al.*, 2019) and post-hoc Multivariate Welch t-test (T^2_w) (Alekseyenko*.*, 2016) which are designed for unbalanced datasets. PERMANOVA and W^*_d tests were

performed to compare host species, site, and the interaction of these two factors, and host species, gut fullness, and their interaction. Sampling year was included as an additional fixed factor. To allow for an interaction between gut fullness and host species, gut fullness was converted into a factor with 2 categories: “full” > 90 %, “partial” ≤ 90 %. Permutations were then constrained by a joint site and gut fullness confounder (strata) and run using 9,999 permutations. *p* values for pairwise T^2_w tests were adjusted using the Benjamini-Hochberg procedure.

To test for differences in alpha diversity between sympatric *Alcolapia* species, Shannon Index values were calculated for each sample’s ASV data using *vegan*. Linear Mixed Models were run using the R package *lme4* v.1.1.28 (Bates *et al.*, 2014). To test for species differences, using sampling year, gut fullness and the interaction between species and gut fullness as additional fixed effects and site as a random effect. A Type I likelihood ratio test was used to test the significance of predictors using the “Anova” function in the *car* v.3.0.12 package (Fox *et al.,* 2012). The “emtrends” function from the *R package emmeans* v.1.7.2 (Lenth, 2022). was used to test for the effect of increased gut fullness on alpha diversity for each species.

Identification of differentially abundant taxa between host species was carried out using *LinDA* (Linear model for Differential Abundance analysis), which uses a mixed effect linear regression modelling approach while correcting for compositional effects (Zhou *et al.*, 2022). This method allows the inclusion of confounding variables to identify the marginal difference in taxa abundance between groups. *LinDA* was run at the level of ASV, family, order and class for the 16S rRNA amplicon dataset. At higher taxonomic groupings, taxa were merged using the “tax_glom” function in *phyloseq* and reads that did not assign to a taxon were discarded. For the 23S plastid dataset, *LinDA* was first run at the level of ASV, but as many taxa only resolved to broad taxonomic levels, ASVs were also merged by level of relatedness using the “tip_glom” function in *phyloseq*. Host species, gut fullness and sampling year were set as the predictors, with an interaction between species and gut fullness. Sampling site was modelled as a random effect. We used the more conservative imputation approach for modelling zeros (Zhou *et al.*, 2022). Samples with less than 1,000 reads and taxa which were present at (≥ 1 reads) in less than 10% of all samples were removed. A quantile winsorization of 0.97 was used to reduce outlier effects, as recommended by Chen *et*

al. (2018) and Zhou *et al.* (2022). To control the false discovery rate, *p* values were corrected using the Benjamini–Hochbergi procedure.

3.2.9 Differences in insect preference between *Alcolapia* species

On observation of COI reads, the majority of non-*Alcolapia* reads were assigned to the Chironomid family (midges). To test for differences between sympatric species preferences towards Chironomidae, a binomial Generalized Linear Mixed Model (GLMM) was run using the *R* package *lme4* (Bates *et al.,* 2014). Presence or absence of Chironomids in a sample was recorded if the relative abundance of Chironomidae reads was above a threshold of 1% (Deagle *et al.,* 2019). Host species, gut fullness, and sampling year were modelled as fixed effects with site as a random effect. The significance of fixed effects was tested using a Type II likelihood ratio test (Fox *et al.,* 2012). Post-hoc Tukey tests were carried out to evaluate pairwise species differences using the *R* package *emmeans* v.1.7.2 (Lenth, 2022).

3.3 Results

Totals of 29,615,916 and 33,239,244 raw reads were obtained for the 23S-16S rRNA and COI MiSeq sequencing runs respectively. After pre-processing and normalisation, the mean numbers of reads per sample and unique ASVs per primer pair were: 13,518 reads per sample and 673 ASVs for plastid 23S rRNA, 9,749 reads per sample and 2,614 ASVs for 16S rRNA, and 31,059 reads per sample and 212 ASVs for COI.

We refer to the gut microbiome in the same way as previous cichlid studies (Franchini *et al.* 2014; Baldo *et al.* 2017; Baldo *et al.* 2019; Härer et al. 2020), which consists of all the non-cyanobacterial taxa that were identified using 16S rRNA. We note that this may contain a combination of gut-resident microbes and externally derived microbes (e.g. from the diet).

We consider the 23S rRNA dataset to be representative of the phototrophs in the *Alcolapia* diet. We note that this excludes any non-cyanobacterial phototrophic bacteria (which may be present in the 16S rRNA dataset). The COI dataset is regarded as the arthropod component of the diet.

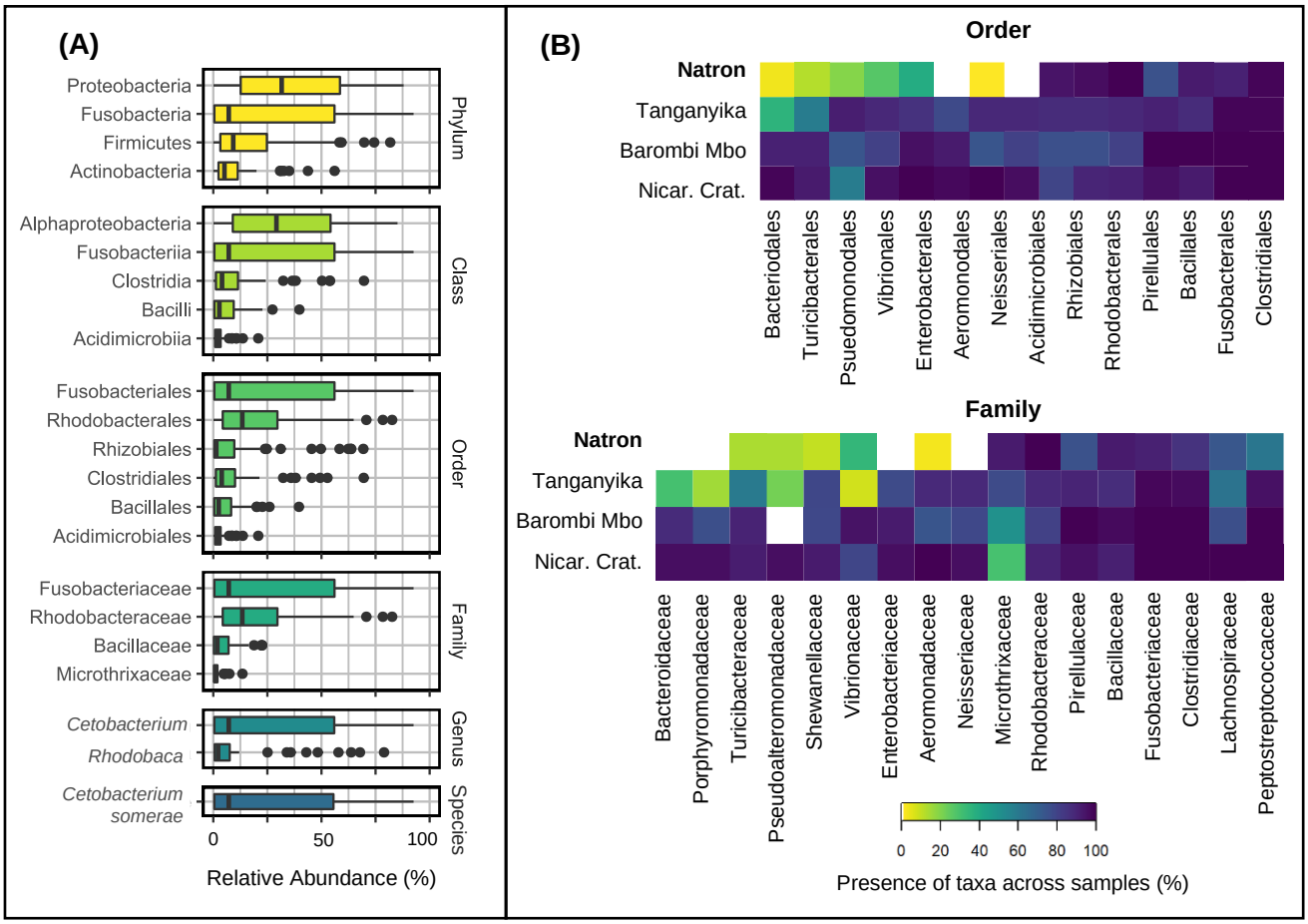


Figure 3.2. The core gut microbiome (16S rRNA) of Lake Natron *Alcolapia* radiation differs to other cichlids. (**A**) Core gut microbial taxa of Lake Natron *Alcolapia* are given at each taxonomic level and are any taxa present in 90% or more of samples with over 10,000 reads (n=56). Core taxa of other cichlids are shown in Figure A3.4. (**B**) Heatmap depicting the differences in the proportion of core microbial orders and families in Natron *Alcolapia* samples compared to other cichlid radiations. Core taxa were any taxa that were present in 90% or more of samples from at least one of these four cichlid radiations (Baldo *et al.,* 2019). 'Nicar. Crat.' refers to cichlids from Nicaragua crater lakes.

3.3.1 The core *Alcolapia* microbiome differs from that of other cichlid radiations

Core microbial taxa were identified in *Alcolapia* at the level of phylum to species (Figure 3.2A). The most abundant phylum in the 16S rRNA dataset were cyanobacterial reads (Figure A3.2), which were removed prior to subsequent analyses. Five phyla accounted for 92% of reads across all samples: Proteobacteria (35%), Fusobacteria (25%), Firmicutes (17%) and Actinobacterota (9%). Three of

System	Genera	Relative abundance (%)	Total species (LPSN)	Halophilic	Alkaliphilic	Phototrophic
Alcolapia	***Cetobacterium***	25.2	2			
	Ca. Chlorothrix	5.2	1			
	Rhodobaculum	4.6	1			
	Terrimicrobium	3.4	1			
	Egibacter	2.5	1			
	IMCC26207	1.7	NA			
	Salipaludibacillus	1.2	5			
	Sarcina	0.7	5			
	Roseibaca	0.7	1			
	Bacillus	0.5	622			
	Ilumatobacter	0.4	5			
	Lactococcus	0.4	26			
	Vibrio	0.3	199			
	Mycobacterium	0.3	263			
	ZOR0006	0.3	NA			
	Natronococcus	0.3	5			
	Alkaliphilus	0.3	12			
	Anaerobacillus	0.3	7			
	Egicoccus	0.3	1			
	Methyloceanibacter	0.3	5	*		
Other cichlids (Baldo *et al.* 2019)	***Cetobacterium***	33.6	2			
	Epulopiscium	2.1	1			
	Plesiomonas	2	1			
	Rhodobacter	1.9	34			
	Porphyromonas	1.3	23			
	Ca. Rhabdochlamydia	1.2	4			
	Clostridium	0.7	326			
	Luteolibacter	0.6	11			
	Bacillus	0.6	625			
	u114	0.6	NA			
	Lawsonia	0.5	1			
	Ca. Xiphinematobacter	0.5	5			
	Pseudomonas	0.4	524			
	Methylosinus	0.4	3			
	Turicibacter	0.3	1			
	Dechloromonas	0.3	8			
	Shewanella	0.3	96			
	Anaerospora	0.2	1			
	Novispirillum	0.2	1			
	Vibrio	0.2	202			

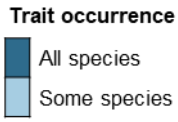


Table 3.3. *Alcolapia* harbour more microbial (16S rRNA) genera that possess adaptations to extreme conditions when compared to other cichlid radiations (Baldo *et al.,* 2019). Included are the top 20 genera by mean read abundance (49% of total reads for *Alcolapia* and 48% for other cichlids). Dark blue = growth characteristic common to all described species in genus, light blue = growth characteristic common to some described species in genus. Taxa common to both *Alcolapia* and other cichlids are highlighted in red. The mean relative abundance (%) and total number of child taxa (species) listed on LPSN for each genus is provided. *Ca.* = *Candidatus*. Genera traits were assigned by referring to described species using the LPSN database (Carte *et al.,* 2020; Table A3.1 and Table A3.2). * *Methyloceanibacter* species are all found in marine environments, but not confirmed as halophilic in laboratory conditions.

these phyla (Proteobacteria, Firmicutes and Fusobacteria) are core cichlid phyla as they were shared by over 90% of samples within each of the other cichlid radiations (Baldo *et al.,* 2019; Figure A3.3). Only 16% of *Alcolapia* samples were assigned reads from the phylum Bacteriodetes despite its presence in 89% or more of samples in each of the other cichlid radiations (Baldo *et al.,* 2019; Figure A3.3).

Alcolapia generally show a lower presence or even complete absence of several core microbial orders and families that are otherwise shared by other cichlid radiations. For instance, the orders Enterobacteriales and Neisseriales (along with their respective families) are entirely absent in all *Alcolapia* samples, despite their presence in 77% or more of samples from each of the other radiations (Baldo *et al.,* 2019). Other taxa show similarly lower presence in *Alcolapia*, such as the orders Alteromonadales (present in 27% of samples) and Vibrionales (39% of samples) and families such as Shewanellaceae (9% of samples) (Figure 3.2B). Moreover, none of the *Alcolapia* core microbial phyla, classes or orders were absent from the core taxa of other cichlids. However, the *Alcolapia* core family Microthrixaceae and genus *Rhodobaca* were exceptions and were only present in 43% and 2.6% of other cichlid samples respectively. Of microbial taxa assigned to the level of species, *Cetobacterium somerae*, a common fish and cichlid gut microbe in the phylum Fusobacteria (Tsuchiya *et al.,* 2008, Baldo *et al.,* 2019), was the only core species in *Alcolapia*, occurring in 91% of the core samples and comprising a mean proportion of 25% of reads (Figure 3.2A).

Of the top 20 most abundant genera found in other cichlids (Baldo *et al.,* 2019), 7 genera contained some species which were either halophilic or alkaliphilic, but none of these genera consisted exclusively of species that possessed these extreme phenotypes (Table 3.3). Conversely, of the 20 most abundant *Alcolapia* genera, 13 contained halophilic species and alkaliphilic species. Moreover, 8 of these most abundant *Alcolapia* genera consisted exclusively of halophilic and alkaliphilic species respectively (Table 3.3).

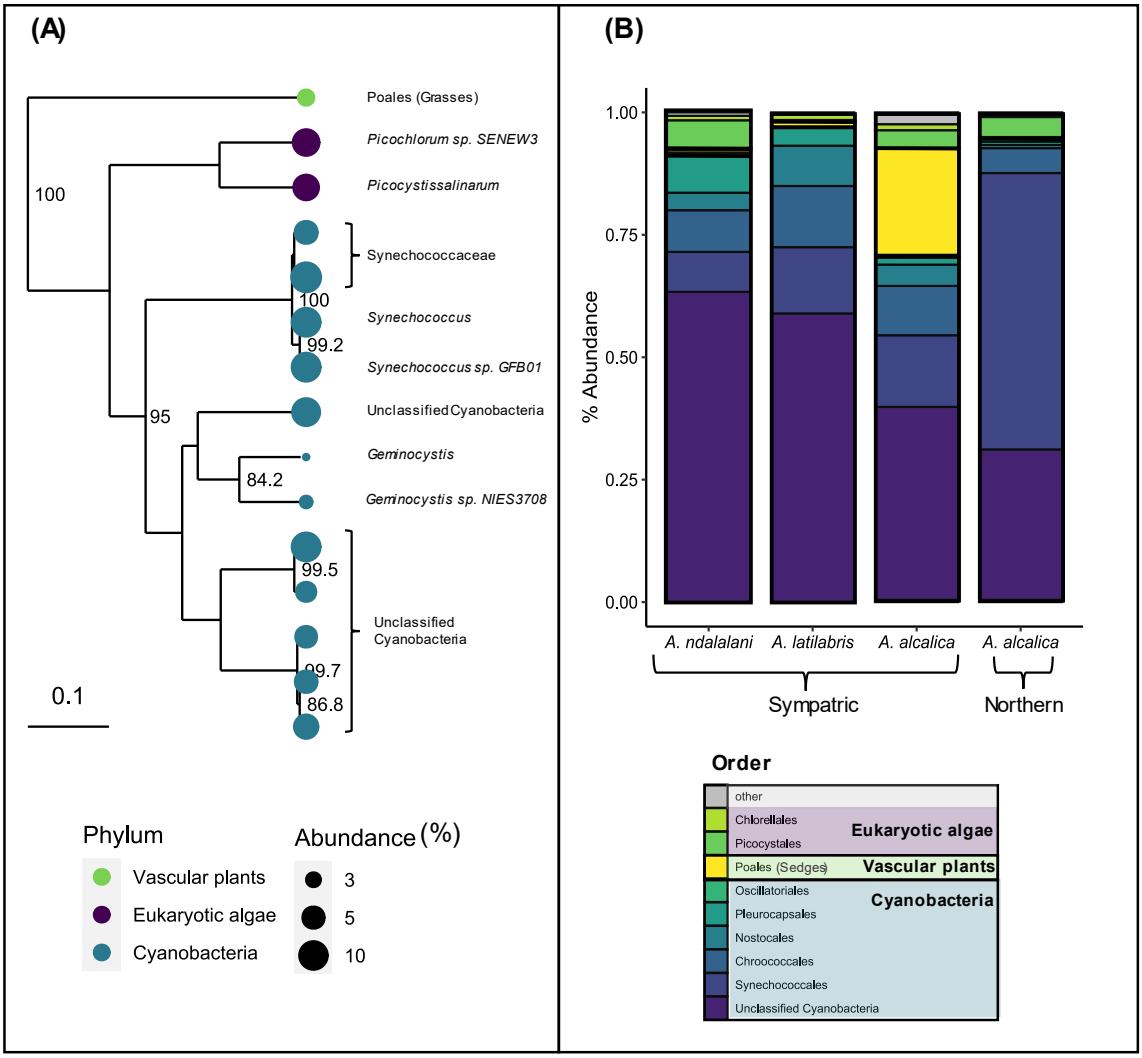


Figure 3.3. Diversity of dietary phototrophic taxa (23S rRNA: algae and plants) in the Lake Natron *Alcolapia* radiation. (**A**) A phylogenetic tree of the top 15 most abundant ASVs in *Alcolapia* from the 23S rRNA plastid amplicon dataset reveals algal diversity. ASV tips are labelled with their highest taxonomic rank after taxonomic assignment and size of ASV tip nodes are proportional to their relative abundance (%) to all reads in the dataset. The phylogenetic tree was generated using *Phangorn* v.2.8.1 with 1000 bootstraps. (**B**) Mean relative abundance (%) of phototrophic dietary taxa in *Alcolapia* species at the level of order. Sympatric *A. alcalica* have a higher proportion of sedges in the diet in comparison to both *A. latilabris* (Log_2FoldChange = -4.3, $p < 0.001$) and *A. ndalalani* (Log_2FoldChange = -4.5, $p < 0.001$). The plot shows the top 9 most abundant orders. “Unclassified Cyanobacteria” are cyanobacterial ASVs that did not classify to order.

3.3.2 *Alcolapia* diets contain high algal diversity, but low invertebrate diversity

Cyanobacteria were the dominant taxa in the 23S rRNA plastid dataset, comprising 85% of reads (Figure 3.3A). However, many of these cyanobacteria were either undescribed or absent in the taxonomic database: 51.3% of cyanobacterial reads were only assigned to the level of class or phylum compared to only 1% of non-cyanobacteria reads (Figure 3.3A). These poorly resolved cyanobacteria ASVs included the first, fifth, seventh and ninth most abundant ASVs (Figure 3.3A). A total of 29% of reads were assigned to the cyanobacterial family Synechococcaceae, which are a diverse and polyphyletic group that are typically unicellular, but can form pseudo-filaments (Dvořák *et al.*, 2014). Other common algal taxa included the eukaryotic and halophilic picoalgae *Picocystis salinarum* and *Picochlorum* sp. *SENEW3* (Krienitz *et al.,* 2012, Foflonker *et al.,* 2016), which comprised 7% and 5.1% of all samples' reads respectively (Figure 3.3A). In addition, 2.1% of reads were assigned to the order Poales (sedges) (Figure 3.3A), with a 100% match to the genus *Cyperus*. This sequence likely derives from the littoral grasses found along the edge of the *Alcolapia* habitat (Figure 3.1C) and may correspond to *Cyperus laevegatus* (Smooth flatsedge), a halophyte known to be present along the shores of Lake Natron (Mwaniki *et al.,* 2019; Kipkemboi, 2016).

The COI amplicon dataset revealed low diversity of arthropod taxa within *Alcolapia* diets (Figure A3.5). Once unassigned reads were removed, 89.8% of reads belonged to the host taxon (Cichlidae) and 9.7% assigned to insects, despite being invertebrate-specific primers. Of reads that were assigned to insects, 95% were from the family Chironomidae (non-biting midges) (Figure A3.5). Chironomidae reads were present in 73% of sympatric samples, with the species *Cladotanytarsus* sp. *KNP8* present in 61% of samples. Other less common taxa recorded included the dragonfly *Paragomphus genei* in 18% of samples, *Macrothrix* sp. *HE-364* (waterfleas) in 7% of samples, and the rotifer *Brachionus dimidiatus* in less than 3% of samples (Figure A3.5).

3.3.3 Sympatric *Alcolapia* species display minor dietary differences

Our results from PERMANOVA (R^2 statistics) and W_d^* tests (p values) found that variation between *Alcolapia* species' dietary phototrophic taxa is largely explained by

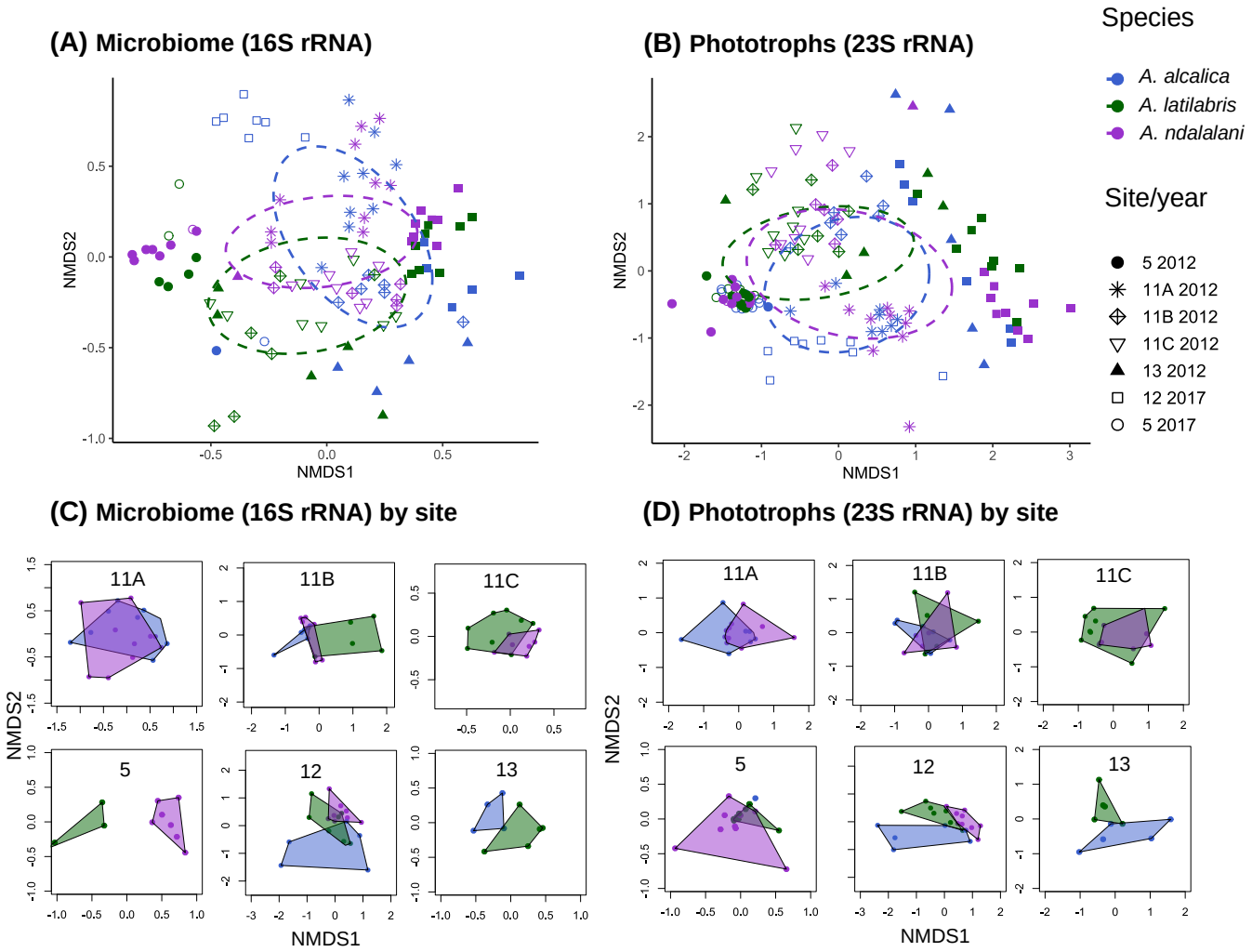


Figure 3.4. Interspecific differences in the microbiome (16S rRNA) and dietary phototrophic taxa (23S rRNA) community composition are small compared to site differences in sympatric *Alcolapia* species. Interspecific differences in the microbiome (**A**) and dietary phototrophic communities (**B**) organised by individual sampling site from specimens collected in 2012. (**C**) and (**D**) show NMDS plots for species at individual sampling site. All plots are Non-metric Multidimensional Scaling (NMDS) using a Bray-Curtis distance matrix of the ASV table. In (**A** and **B**), ellipses show the 95% confidence interval of the centroid for each species based on the standard errors of average axis scores. Plots using the generalised UniFrac are provided in Figure A3.6.

site differences when multiple distance metrics are considered (Figure 3.4; Table A3.2; Figure A.3.6). For instance, using the Bray-Curtis distance metric, site explained most of the variation (PERMANOVA R^2 = 0.57; W_d^*stat = 35.8, p < 0.001) while overall species differences were significant but minor (R^2 = 0.01; W_d^*stat = 2.12, p = 0.002; Table A3.2). However, the interaction between site and species explained a greater proportion of the variance (R^2 = 0.05; W_d^*stat = 25.2, p < 0.002; Table A3.2), indicating that the extent of species differences varies depending on their site (Figure 3.4). Small

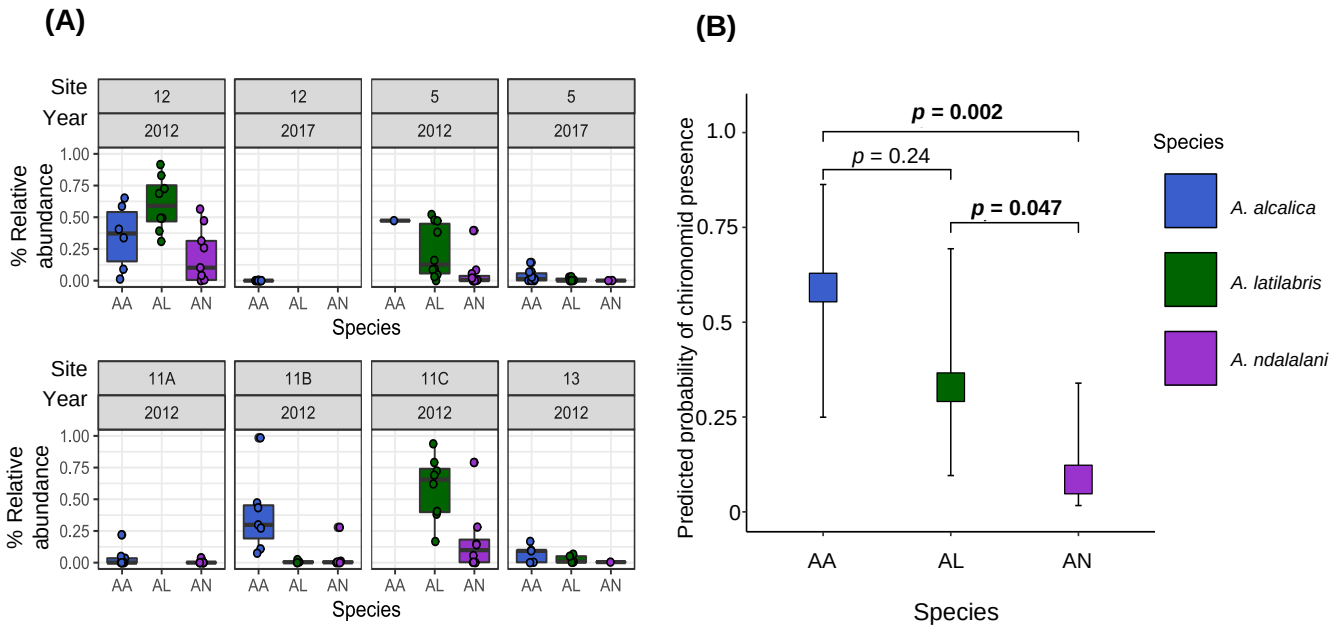


Figure 3.5. Interspecific differences in gut midge (Chironomidae) presence among sympatric Lake Natron *Alcolapia* species. **(A)** The proportion of total metazoan reads assigning to Chironomid (midge fly) varies across sympatric species at different sites and years. **(B)** The predicted probability of Chironomid read presence (at > 1% of total reads) in species from sympatric sites. The estimated marginal means and 95% Wald confidence intervals are taken from the *post-hoc* test from a binomial GLMM, with Tukey adjustment to allow for multiple comparisons. Significant *p*-values are highlighted in bold.

but significant differences in community composition depending on gut fullness were recorded (e.g. Bray-Curtis, R^2 = 0.01, T_W^2 = 35.8, p = 0.001). In pairwise comparisons between species, the most consistent differences in phototrophic dietary composition were between *A. alcalica* and *A. latilabris* (e.g. Bray-Curtis, T^2_w = 2.37, p = 0.006; Table 3.4). Pairwise comparisons between *A. ndalalani* and the other species were only significant using the unweighted UniFrac metrics (vs. *A. alcalica*: T^2_w= 2.00, p = 0.005; vs. *A. latilabris*: T^2_w = 2.33, p = 0.005), suggesting that any species differences that may exist are likely driven by rare and phylogenetically dissimilar ASVs (Table 3.4). There were no significant differences in dietary alpha diversity (Shannon's Index) between sympatric *Alcolapia* species (X^2 = 2.47, df = 2, p= 0.29). Gut fullness levels had no significant effect on alpha diversity (X^2 = 0.05, p= 0.83). However, samples collected in 2017 had significantly lower diversity than 2012 samples (X^2 = 11.2, df = 1, p < 0.001; Figure A3.7).

Differential abundance analysis found evidence that sympatric *A. alcalica* feed significantly more on Poales (sedges) compared to both benthic species, and that they had significantly more ASVs assigned to the order Poales compared to both *A. latilabris* (LinDA, Log2FoldChange = -4.3, $p < 0.001$) and *A. ndalalani* (LinDA, Log2FoldChange = -4.5, $p < 0.001$; Figure A3.8). *Alcolapia latilabris* and *A. ndalalani* were also enriched in some cyanobacteria ASVs in comparison to *A. alcalica* (Figure A3.8).

Alcolapia species were found to vary in their presence of chironomids (Binomial GLMM, $X^2 = 12.0$, $df = 2$, $p < 0.003$; Figure 3.5). Post-hoc tests found a lower occurrence in *A. ndalalani* than in *A. alcalica* (post-hoc Tukey, *z*.ratio = 3.38, $p = 0.002$). *A. latilabris* also had a higher occurrence of chironomids than *A. ndalalani*, but it was only marginally significantly (*post-hoc* Tukey, *z*.ratio = 2.37, $p = 0.047$). There were no statistically significant differences between *A. alcalica* and *A. latilabris* (post-hoc Tukey, *z*.ratio = 1.62, $p = 0.236$; Figure 3.5B). Samples from 2012 had a significantly higher presence of midges compared to 2017 samples ($X^2 = 9.09$, $df = 1$, $p = 0.003$; Figure 3.5A), but there were no significant differences in chironomid presence depending on gut fullness ($X^2 = 0.87$, $df = 1$, $p = 0.35$).

3.3.4 Sympatric *Alcolapia* species have different gut microbiota

Similar to the phototrophic algae data (Section 3.3.3), our results from PERMANOVA (R^2 statistics) and W_d^* tests (*p* values) found that variation in the *Alcolapia* microbiome's (16S rRNS) composition is largely explained by site differences when multiple distance metrics are considered (Figure 3.5, Table A3.3, Figure A3.9). For instance, using the Bray-Curtis distance metric, site explained a large amount of the variation (PERMANOVA $R^2 = 0.31$; W_d^*stat = 9.02, $p < 0.001$), whereas overall species differences were significant but minor ($R^2 = 0.06$; W_d^*stat = 2.12, $p < 0.001$). However, the interaction between site and species explained a greater proportion of the variance ($R^2 = 0.11$; W_d^*: stat = 5.80, $p = 0.015$, Table A3.3), again suggesting that the extent of species differences varies depending on site (Figure 3.5). Gut fullness levels also slightly affected the microbiome composition ($R^2 = 0.02$, $T_W^2 = 3.81$, $p < 0.001$).

Post-hoc T^2_w pairwise tests identified species differences across all distance metrics (Bray-Curtis, unweighted UniFrac, and generalized UniFrac) (Table 3.4). For instance,

		Microbiome (16S rRNA)						Phototrophs (23S rRNA)					
		Bray-Curtis		Unweighted Unifrac		Generalised Unifrac		Bray-Curtis		Unweighted Unifrac		Generalised Unifrac	
		T^2_w	p	T^2_w	p	T^2_w	p	T^2_w	p	T^2_w	p	T^2_w	p
A. alcalica	*A. latilabris*	2.98	**<0.001**	2.99	**0.004**	3.24	**0.002**	2.38	**0.002**	2.49	**0.001**	2.42	**0.002**
A. alcalica	*A. ndalalani*	2.82	**0.002**	2.27	**0.012**	2.82	**0.004**	2.10	0.119	2.00	**0.005**	1.25	0.367
A. latilabris	*A. ndalalani*	3.45	**<0.001**	2.21	**0.014**	6.06	**<0.001**	1.91	0.119	2.33	**0.005**	2.17	0.199

Table 3.4. Sympatric *Alcolapia* species have more consistent differences in their gut microbiome (16S rRNA) compared to their dietary phototroph (23S rRNA) composition. Post-hoc tests were carried out on three distance measures (Bray-Curtis, unweighted and Generalised Unifrac) of ASV tables using T^2_w tests based on Welch's MANOVA). *p* values and test statistics (T^2_w) were generated for each term with a joint site and gut fullness cofounder (strata) and 9,999 permutations. *p* values were adjusted to control for false discovery rate using the Benjamini-Hochberg procedure. The samples sizes for each species for the microbiome post-hoc tests was: *A. alcalica* = 26, *A. latilabris* = 35, *A. ndalalani* = 39, and for the phototrophic taxa post-hoc test: *A. alcalica* = 36, *A. latilabris*= 42, *A. ndalalani* = 44.

using the Bray-Curtis distance metric, there was a significant difference in the gut microbial communities of *A. alcalica* and the benthic cichlid species (*A. alcalica* vs *A. latilabris*: $T^2_w = 2.98$, $p < 0.001$; *A. alcalica* vs *A. ndalalani*: $T^2_w = 2.82$, $p = 0.002$) and between *A. latilabris* and *A. ndalalani* ($T^2_w = 3.45$, $p = <0.001$; Table 3.4). Alpha diversity (Shannon Index) differences between species were non-significant ($X^2 = 2.37$ $df = 2$, $p = 0.11$), but there was a significant interaction between host species and gut fullness level ($X^2 = 3.79$, $df = 2$, $p = 0.029$). Alpha diversity was significantly greater with increased gut fullness in *A. latilabris* (*t*.ratio = 2.05, $df = 97.6$, $p = 0.018$), but marginally non-significant in *A. alcalica* (*t*.ratio = 2.18, $df = 96.2$, $p = 0.051$) and not significant in *A. ndalalani* (*t*.ratio = -0.18, $df = 97.3$, $p = 0.77$). The effect of year was not significant ($X^2 = 1.61$, $df = 1$, $p = 0.08$).

We identified several taxa which were differentially abundant between the gut microbiomes of sympatric *Alcolapia* at the level of class, order, family and ASV (Figure A3.9). Notably, *A. alcalica* and *A. latilabris* were enriched in the class Fusobacteria compared to *A. ndalalani*. This class includes the core cichlid genus *Cetobacterium*,

which made up the top three and four most differentially abundant ASVs found in *A. alcalica-A. ndalalani* and *A. latilabris-A.ndalalani* comparisons respectively (Figure A3.9). *Alcolapia ndalalani* were also enriched in the classes Acidimicrobia and Alphaproteobacteria in comparison to the other species, and in Verrucomicrobiae compared to *A. latilabris* (Figure A3.9). Furthermore, compared to *A. alcalica*, *A. ndalalani* had a higher relative abundance of the class Chloroflexia (non-sulphur green bacteria) and some lower taxonomic ranks included in this clade such as the photosynthetic *Candidatus* Chlorothrixa (Klappenbach and Pierson, 2004). Conversely, *A. latilabris* and *A. alcalica* showed no differentially abundant classes (Figure A3.9).

3.4 Discussion

3.4.1 Main findings

Dietary niche divergence is a primary driver of ecological speciation, especially among sympatric populations (Turelli *et al.*, 2001; Rundle and Nosil, 2005; Faria *et al.*, 2014). Increasingly, the gut microbiome has gained interest in the speciation field as it can be a signature of diet and habitat niche segregation, as well as a potential source of variation for selection (Shropshire and Bordenstein, 2016; Miller *et al.*, 2021). We characterised the cryptic diversity in both the *Alcolapia* diet and gut microbiota using a metabarcoding approach, with a combination of universal plastid, arthropod and bacterial primers.

We discovered high diversity in dietary algal taxa at some sites, but many of the most abundant sequences only resolved to broad taxonomic categories. *Alcolapia* were found to lack some core microbial taxa found in other cichlids but were enriched in extremophile taxa, suggesting that the hostile conditions may influence gut microbial communities. Collection site was a better predictor of diet and microbiome, but sympatric *Alcolapia* species exhibited minor differences, with some differentially abundant taxa identified for each species.

3.4.2 The *Alcolapia* microbiome

Our findings indicate that *Alcolapia* lack some of the core taxa found in other cichlids. This is perhaps unexpected given that some of these core microbial taxa are shared

between cichlids separated across continents and over large phylogenetic distances. While the three most abundant phyla in *Alcolapia* and other cichlids were the same (Proteobacteria, Fusobacteria and Firmicutes), the relative abundance of Proteobacteria was much higher in *Alcolapia* (Baldo *et al.*, 2019). The high levels of Proteobacteria in *Alcolapia* guts may be environmentally derived from the soda lake environment. For instance, Kambura *et al.* (2016) found that Proteobacteria were the most abundant microbial phylum in Lake Magadi hot springs.

The gut microbiota is composed of both autochthonous (gut residents) and allochthonous (non-residents) microbes, which are difficult to differentiate. Studies in cichlids and other fish have typically found that allochthnonous sources comprise a minor component of the microbiome (Bolnick *et al.*, 2014; Giatsis *et al.*, 2015; Baldo *et al.*, 2017; Kim *et al.*, 2021). However, the greater abundance of halotolerant and alkaliphilic-associated microbial genera in *Alcolapia* guts suggests that some of these species may be allochthonous. While *Alcolapia* have evolved physiological innovations to avoid alkaline water entering the stomach, *A. grahami* was also found to have an increasingly alkaline gradient along the gastrointestinal tract, with a pH in the hindgut of 8.12 and 8.42 at the rectum (Bergman *et al.*, 2003). Some alkaliphilic microbes may therefore transiently or even permanently reside in parts of the gut. Nonetheless, this absence of some core taxa and presence of more extremophile taxa in *Alcolapia* relative to other cichlids suggests that the soda lake conditions have a major effect on the internal gut microbiome composition.

Many members of the core cichlid taxa are mesophilic and are common as both vertebrate commensal gut microbes and pathogens (Morse, 2011; Patel *et al.*, 2014; Baldo *et al.,* 2019). Since the fish gut microbiome is partially derived from the environment (Ley *et al.*, 2008), a possible explanation for an absence of some core cichlid taxa could be that some of these bacteria may be unable to colonise the *Alcolapia* gut due to being unable to survive in the extreme conditions outside the host, causing these taxa to become secondarily lost (Wong and Rawls, 2012; Kim *et al.*, 2021). Additionally, the presence of extremophile-associated microbes may represent novel colonisation of the gut by microbial taxa originating from the soda lake environment, which may in turn displace other core taxa.

The lower number of core microbial taxa in *Alcolapia* may also result from the lack of trophic diversity in a diet largely consisting of cyanobacteria. Specialist or simplified diets potentially cause a reduction in gut microbiota diversity compared to hosts with more generalist diets (Dill-Mcfarland *et al.*, 2016; Amillano-Cisneros *et al.*, 2022). Comparisons with closely related sister clades (e.g. other *Oreochromis* species) and laboratory manipulations of diet and environmental conditions may be required to disentangle the processes modulating these microbial communities.

3.4.3 Cryptic diversity in the *Alcolapia* diet

Consistent with previous findings, we found that the algal DNA in the *Alcolapia* gut is dominated by cyanobacterial reads (Coe, 1966; Ford *et al.*, 2016). However, the presence of multiple cyanobacterial orders revealed a large degree of cryptic diversity which was not detectable using traditional gut content analysis (Ford *et al.,* 2016), echoing the findings from phylogenetic studies of soda lakes (Mikhodyuk *et al.,* 2008; Krienitz and Schagerl, 2016). Much of this cyanobacterial diversity remains to be sequenced or characterised, as we found a significant proportion of cyanobacterial reads were only assigned to the level of phylum or class. Some of these unknown ASVs are likely to be variants the same species as they are genetically similar. The cyanobacterial orders Synechoccales, Chroococcales, Nostocales, Pleurocapsales and Oscillatoriales were present, which have previously been documented in Lake Magadi (Jones and Grant, 1999; Edwin *et al.,* 2019). These diverse orders include both planktonic species, but also filamentous species that are potentially targeted by *Alcolapia* in benthic mats (Komárek and Johansen, 2015). The lack of *Arthrospira fusiformis*, which is a highly abundant character cyanobacterial species from East African soda lakes was unexpected, as it has been reported to comprise a major component of the diet of *A. grahami* (Coe, 1966; Vareschi and Vareschi, 1979; Nonga *et al.*, 2011; Schagerl *et al.*, 2015; Krienitz and Schagerl, 2016). However, Tebbs *et al.* (2013) do note that *Arthrospira* is limited in Lake Natron and as a planktonic species it may not be targeted by *Alcolapia.* On the other hand, *A. grahami* introduced to Lake Nakuru reportedly switched to filter feeding of planktonic *Arthrospira* (Vareschi and Vareschi, 1979). The presence of eukaryotic algae such as the halophilic picoalgal species *Picocystis* and *Picochlorum* indicates that *Alcolcapia* may also feed on other algal taxa (Foflonker *et al.*, 2016; Krienitz *et al.*, 2019). It is unclear to what extent

these various algal taxa are intentionally targeted by *Alcolapia* or are ingested indiscriminately from the water column, especially as *Alcolapia* are known to drink at high rates (Wood *et al.*, 2002; Bergman *et al.*, 2003). In addition, the relative abundance of amplicon sequences may only have a weak quantitative relationship with biomass, making it difficult to infer the absolute contributions of these different dietary taxa (Deagle *et al.*, 2019; Lamb *et al.*, 2019).

Our findings using macroinvertebrate primers showed there is a limited diversity of invertebrate fauna in *Alcolapia* diets (Melack, 1996; Mengistou, 2016). There is a lack of studies on the macroinvertebrate assemblages of Lake Natron (Mengistou, 2016), but consistent with studies of other East African soda lakes, the most common invertebrates were chironomids (non-biting midges), which can form a significant proportion of the insect biomass (Mengistou, 2016). For instance, Harrison (2002) recorded significant numbers of *Cladotanytarsus* in Ethiopian soda lakes, a genus which were also found in over half of the specimens here. Visual analysis of stomach contents using specimens from the same collection trip as used here found that insect and arthropods comprised a small proportion of the diet (Ford *et al.*, 2016). Given the high insect biomass, insectivory may be considered as a viable niche strategy during certain priods within these habitats. Indeed, *A. grahami* that were introduced to Lake Nakuru dramatically reduced mosquito abundance (Matagi, 2011). The higher proportion of chironomid reads which we identified in specimens collected in 2012 compared to those from 2017 may represent seasonal variation in chironomid abundance as they were collected at different times of year (January and August), which are associated with different rainfall and temperature patterns (Vincens and Casanova, 1987). Temporal fluctuations in insect abundance may therefore preclude the permanent niche specialisation towards insectivory.

3.4.4 Differences in *Alcolapia* species' diets

Our results largely corroborate with the previous results based on stable isotopes and stomach contents, which find that Natron *Alcolapia* have considerable overlap in their diets (Ford *et al.*, 2016). We found a higher presence of sedge DNA in *A. alcalica* guts, confirming that they feed on more vascular plants compared to the other *Alcolapia* species (Ford *et al.*, 2016). However, a lack of differences in dietary algae between both benthic species, *A. latilabris* and *A. ndalalani,* suggests they largely feed on the

same algal species. This phenomenon of high dietary overlap despite differences in trophic morphology is found in other sympatric cichlid species such as rocky shore mbuna from Lake Malawi and Midas cichlids from Nicaraguan crater lakes (Bootsma *et al.*, 1996; Genner *et al.,* 1999a, 1999b; Stauffer Jr and McKaye, 2002; Martin and Genner, 2009). One potential explanation for this paradox is that specialist trophic morphologies only become functionally adaptive in periods of resource scarcity (Liem, 1980; McKaye and Marsh, 1983; Bootsma *et al.*, 1996; Galvez *et al.*, 2022). Here, the degree of dietary overlap varied spatially and temporally, lending some support to this facultative specialist hypothesis. Another explanation for this phenomenon could be the relationship of trophic morphology to foraging behaviour. Differences in their mouth morphology, stomach contents and observations of their feeding behaviour in the wild reveal that *A. latilabris* feeds largely by scraping algae from rocks whereas *A. ndalalani* browses, so while they feed largely on the same trophic sources, they may use a different foraging mode (Yamaoka, 1991; Ford *et al.*, 2016).

In interspecific comparisons, we also found a higher proportion of chironomid DNA in both *A. alcalica* and *A. latilabris* guts than in *A. ndalalani* guts. While we cannot quantitatively estimate what proportion of the overall diet that is comprised of chironomids, previous analyses of stomach contents found that insect parts comprised a very minor component of the diet (Ford *et al.*, 2016). *A. alcalica* has a more generalist mouth morphology that may be more suited to predation of surface insects compared to the two benthic species (Seegers and Tichy, 1999; Ford *et al.*, 2015). *A. alcalica* at some multispecies sites are more abundant along the littoral margins of the spring, which may also be the preferred habitat for dipterous flies (G.F. Turner personal communication, 16th July 2022; Figure 3.1C). The lower proportion of chironomid DNA in *A. ndalalani* relative to *A. latilabris* may be related to their foraging mode. With a wider mouth and hypertrophied lips, *A. latilabris* feeds by rasping at rocks and intakes more detritus which may inadvertently contain insect parts (Ford *et al.*, 2016).

3.4.5 The *Alcolapia* microbiome and speciation

Our findings suggest that the *Alcolapia* gut microbiome composition is determined primarily by habitat variation with a smaller effect of differing host species. This strong location effect is seen in many taxa and particularly in fish (Ghanbari *et al.*, 2015; Smith *et al.*, 2015; Kim *et al.*, 2021). While we could not detect a difference in alpha diversity

between *Alcolapia* species, beta diversity analysis found differences among all species using multiple metrics. The greater consistency of species differences in the gut microbiome relative to the diet may reflect the more stable nature of the microbiome over time.

While we were unable to separate autochthonous and allochthonous microbes in this study, some taxa were strongly associated with either the gut or external environment. For instance, *A. ndalalani* had significantly lower levels of *Cetobacterium* compared to the other species. *Cetobacterium somerae* is a fermenter of carbohydrates and peptides (Tsuchiya *et al.,* 2008) and is a known commensal gut microbe of cichlids, often making up the majority of the gut microbiome (Baldo *et al.*, 2019). It may have functions in promoting fish health such as through generating vitamin B12 or regulating aerobic metabolism in hypoxic conditions (Tsuchiya *et al.,* 2008; Bhute *et al.*, 2020). Differences between *Alcolapia* species were also observed in the abundance of microbial taxa associated with the environment. For instance, *A. ndalalani* had a greater abundance of reads from the families Rhodobacteracea (non-sulphur purple bacteria) and Chloroflexaceae (non-sulphur green bacteria), both of which contain many phototrophic species such as *Candidatus* Chlorothrix and *Roseibaca* (Klappenbach and Pierson, 2004; Labrenz *et al.*, 2009; Pujalte *et al.*, 2014; Baldo *et al.*, 2017; Madigan *et al.,* 2017). It is unknown to what extent these taxa are directly targeted by *Alcolapia* and how they may contribute to the diet, although non-sulphur purple bacteria have previously been shown to have beneficial nutritional value to *Oreochromis* (Banerjee *et al.*, 2000; Delamare-Debouteville *et al.*, 2019).

The association of the *Alcolapia* microbiome composition with host species identity suggests there could be a potential phylosymbiotic relationship (Brooks *et al.*, 2016; Kohl, 2020). *A. latilabris* and *A. ndalalani* may be expected to have more similar microbiomes compared to *A. alcalica* as they group as sister species in phylogenomic analyses, although extensive gene flow between all three species adds additional noise to these relationships (Ford *et al.*, 2015). Given the small number of taxa and the recent speciation in this adaptive radiation, there is too little power to explicitly test whether phylosymbiosis is occurring in this system.

3.4.6 The role of ecological processes in *Alcolapia* speciation

Ecological opportunity is considered an important predictor of adaptive radiation (Wagner *et al.,* 2012; Wellborn and Langerhans, 2015). Depauperate habitats may have high ecological opportunity due to the availability of unexploited niches. This makes them primed for adaptive radiations, as incipient species evolve to occupy vacant niches (Wellborn and Langerhans, 2015). In addition, productive habitats are predictors of species richness (Evans *et al.,* 2005; Wagner *et al.*, 2012) and while lacking in diversity of flora and fauna, East African soda lakes are among the most productive ecosystems on earth (Oduor and Schagerl, 2007). In this regard, extreme environments are paradoxical in their potential for adaptive radiation because while they are depauperate, the hostile conditions may also prevent the colonisation of other diverse taxa that can increase trophic complexity, thereby reducing the availability of different niches to exploit. In Natron, this unusual combination of low niche complexity but high productivity and a lack of competition to fill vacant niches may in part explain how *Alcolapia* have managed to rapidly evolve three species with divergent trophic morphology along a remarkably narrow trophic axis (Ford *et al.*, 2016). Our findings of host species differences in gut microbiome composition, despite *Alcolapia* having a lower number of core taxa compared to other radiations, suggests that this narrow niche differentiation may also extend to the microbial communities in the gut. Laboratory manipulations could be used to disentangle the extent to which these species differences are linked to diet and habitat or associated to host-specific factors such as physiology.

CHAPTER 4: DEMOGRAPHIC MODELLING REVEALS EARLY DIVERGENCE AND HISTORICAL HYBRIDISATION IN THE *ALCOLAPIA* RADIATION

ABSTRACT

Disentangling past patterns of divergence and gene flow is crucial to understanding how speciation occurs over time. Young and small species flocks from isolated regions are valuable targets for studying these past processes, as the number of possible scenarios is more limited compared to older and more complex systems. Here, we used demographic modelling to investigate historical processes of speciation in soda lake cichlids (*Oreochromis* (*Alcolapia*) spp.), a small species flock of cichlid fish endemic to the extreme and isolated lakes Natron (3 species) and Magadi (1 species) in East Africa. In Natron, the three species exist sympatrically in small and shallow habitats with ongoing interspecific gene flow. Previous approaches have been unable to fully resolve the species' phylogenetic relationships. Moreover, it was unclear whether the Natron species diverged before or after the lakes split, which has implications for our knowledge of the rapid and sympatric nature of their speciation. We used whole genome re-sequenced data and a published RADseq dataset to quantify the patterns of gene flow between populations and to infer the demographic history of the radiation. We found that rates of introgression varied between species and sub-populations, with evidence of asymmetric migration between species pairs, presenting a more complex pattern of gene flow in the radiation. Demographic models supported the paraphyly of the Natron radiation, with the common ancestor estimated to exist before the lake split, indicating that all three Natron species did not emerge sympatrically in the current extreme soda conditions. Our findings of an earlier divergence time and high rates of gene flow between ancestral *Alcolapia* lineages highlight the importance of exploring past processes when studying cases of recent speciation in a sympatric setting.

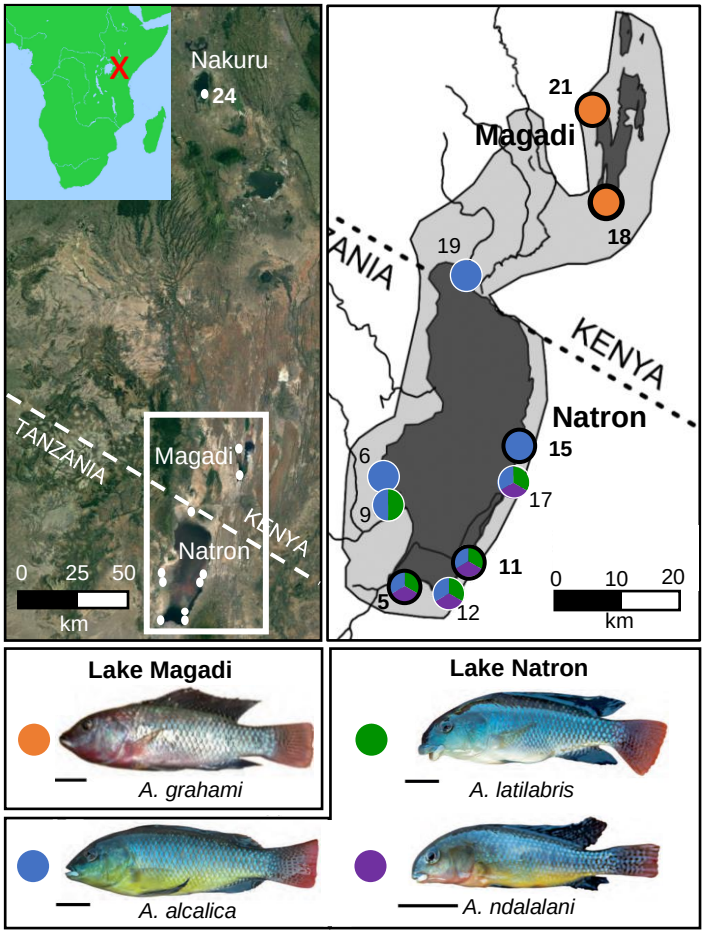
4.1 Introduction

Speciation in the presence of gene flow is common among recently diverging lineages with overlapping distributions (Coyne and Orr, 2004; Nosil, 2008). Gene flow can also occur between taxa that have diverged in geographic isolation if they come into secondary contact before reproductive isolation is complete (Feder *et al.*, 2013). In both these contexts, gene flow can impede or even reverse the speciation process as recombination breaks down loci associated with ecological selection or assortative mating, causing populations to homogenise (Slatkin, 1987; Feder *et al.*, 2012; Foote, 2018). However, gene flow may also promote speciation through reinforcement by driving reproductive character displacement (Butlin, 1987; Servedio and Noor, 2003; Pfennig and Pfennig, 2009). Hybridisation may also transfer adaptive genes via introgression or even generate novel adaptive phenotypes, potentially leading to hybrid speciation (Rieseberg *et al.*, 1999; Mallet, 2008; Abbott *et al.*, 2013). Quantifying gene flow between incipient species allows inferences of the extent of reproductive isolation between species and is a key step before the role of hybridisation in speciation can be evaluated (Dieckmann and Doebeli., 2012; Stankowski and Ravinet, 2021; Bolnick *et al.*, 2023).

In many contexts, speciation is not strictly linear in its chronology, and alterations in the rate of gene flow may cause populations to either converge or diverge over time. These earlier processes can strongly influence how speciation proceeds, with fluctuations in population connectivity, changes in population size and ancestral polymorphisms having a significant effect on subsequent speciation (Feder *et al.*, 2012; Guerrero and Hahn, 2017; Marques *et al.*, 2019). While characterising reproductive barriers between lineages can provide a glimpse at the current isolating mechanisms, it is unable to offer insights into all the past processes underpinning divergence. Therefore, a holistic framework that incorporates analyses of the species' historical evolution is required to create a complete picture of speciation (Smadja and Butlin, 2011; Butlin *et al.*, 2012).

Past patterns of speciation can be deduced in multiple ways, including by studying the biogeography and distribution of current populations or by analysing fossil evidence

and genomic datasets (Barton, 1988; Barraclough and Nee, 2001; Coyne and Orr, 2004).



Lake	Site	Species/ morph	No. samples	
			RAD	WG
Natron	5	*A. alcalica*	4	
		A. latilabris	4	1
		A. ndalalani	4	1
	11	*A. alcalica*	4	3
		A. latilabris	4	1
		A. ndalalani	4	1
	12	*A. alcalica*	8	
		A. latilabris	4	
		A. ndalalani	4	
	17	*A. alcalica**	4	
		*A. latilabris**	4	
		*A. ndalalani**	4	
	9	*A. alcalica*	4	
		A. latilabris	4	
	15A	*A. alcalica*	4	1
	15B	*A. alcalica* (upturned)	4	2
	6	*A. alcalica*	4	
	19	*A. alcalica*	4	
Magadi	**18**	*A. grahami*	4	1
	21	*A. grahami*	4	1
Nakuru	**24***	*A. grahami*	4	2

Figure 4.1. Origin of the *Alcolapia* spp. genomic samples used in this study. Top left panel: Map showing the distribution of *Alcolapia* in the East African rift valley with endemic Lake Natron harbouring three species (bottom left panel: *A. alcalica*, *A. latilabris* and *A. ndalalani*) and Lake Magadi harbouring a single species (*A. grahami*). An additional location (site 24) at Lake Nakuru contains *A. grahami* that were introduced from Magadi in the 1950's. Middle top panel: endemic sites sampled with genomic data included for this study. Coloured circles correspond to species present in the bottom left panel (scale bar = 1cm). Sites outlined in bold contain additional whole genome (WG) re-sequenced samples. Dark grey areas denote the current extent of the lake, whereas the light grey area shows the approximate maximum extent of Palaeolake Orolonga (from ~700 Ka) (Williamson *et al.,* 1993). Right table: the number of RADseq (Ford *et al.,* 2015) and whole genome re-sequenced (WG) samples by site and species. Photos, site information and schematic map are adapted from Ford *et al*. (2015).

Small and isolated adaptive radiations make for ideal candidates for studying these past processes as they offer more tractable systems where confounding effects and

alternate speciation scenarios are more easily ruled out (Coyne and Orr, 2004). However, even in these simpler systems, inferences can be a challenge if the radiation has been rapid and current populations overlap in their distributions (Richards *et al.,* 2019; Olave and Meyer, 2020). In recent years, advances in high-throughput sequencing have enabled the fast and affordable retrieval of genomic data of multiple individuals (Feder *et al.*, 2012; Foote, 2018; Salzburger, 2018). Genomic data contains a plethora of information that can enable inferences about a taxon's past evolutionary and demographic history. These approaches move beyond simple dichotomous trees and incorporate additional parameters such as gene flow and ancestral population sizes (Reich *et al.*, 2010; Frantz *et al.*, 2013; Beichman *et al.*, 2018).

As a small and recent radiation, soda lake cichlid fishes (*Oreochromis* (*Alcolapia*) spp.) make for an ideal system to investigate the process of speciation with gene flow. Endemic to soda lakes Natron and Magadi in the East African rift valley (Seegers and Tichy, 1999), these cichlids inhabit some of the most hostile aquatic habitats of any vertebrate and experience high pH (~10), temperature (< 42°C), salinity, UV light and oxygen levels (Coe, 1966; Pörtner *et al.*, 2010; Johannsson *et al.*, 2014). They have long generated research interest for their physiological adaptations to these extreme soda lake conditions (Trewavas, 1983; Narahara *et al.*, 1996; Walsh *et al.*, 2001; Wood *et al.*, 2012, 2016; White *et al.*, 2020, 2022). Despite these harsh conditions, *Alcolapia* has formed a small adaptive radiation including four described species and a range of intraspecific morphological variants both in Natron and Magadi (Seegers *et al.*, 2001; Ford *et al.*, 2015; Kavembe *et al.*, 2016). This adaptive radiation has occurred extremely rapidly, with the radiation estimated to be less than 10 ka (Tichy and Seegers, 1999; Ford *et al.*, 2015, 2019; Kavembe *et al.*, 2015).

Three species are described from Lake Natron (*Oreochromis* (*Alcolapia*) *alcalica*, *Oreochromis* (*Alcolapia*) *ndalalani* and *Oreochromis* (*Alcolapia*) *latilabris*), while a single species (*Oreochromis* (*Alcolapia*) *grahami*) is described from the nearby Lake Magadi and its satellite lake Little Magadi (Coe, 1966; Seegers *et al.*, 2001; Ford *et al.*, 2019). Populations of *A. grahami* have also been introduced to Lake Nakuru, Kenya (Figure 4.1). Significant hybridisation between sympatric species suggests that species are still at the early stages of divergence (Ford *et al.,* 2015; Chapter 2).

Phylogenetic analyses indicate that within the Natron radiation, *A. ndalalani* and *A. latilabris* are sister species, whereas *A. alcalica* comprises two separate lineages: a northern lineage where it occurs allopatrically and a southern lineage which co-occurs with *A. ndalalani* and *A. latilabris* (Ford *et al.*, 2015; Figure 4.1). Among sympatric Natron species, there are relatively high levels of interspecific gene flow, indicating incomplete reproductive isolation (Tichy and Seegers, 1999; Zaccara *et al.*, 2014; Ford *et al.*, 2015). Compared to most other freshwater lakes inhabited by cichlids, Natron and Magadi have an atypical structure. The lake interiors are dominated by sodium hydrogencarbonate precipitates (trona) for most of the year, with *Alcolapia* inhabiting volcanic spring sites, streams and lagoons on the edges of the lake (Coe, 1966, 1969; Seegers *et al.*, 2001). Particular sites contain different species and morph compositions with varying levels of connectivity and gene flow (Seegers and Tichy, 1999; Ford *et al.*, 2015). In addition, there is intraspecific variation in colour and morphology, and the relative frequency of the different species can vary between sites (Seegers and Tichy, 1999; Tichy and Seegers, 1999; Ford, 2015). Fluctuations in the lake habitat structure (e.g. from major flooding events) may also cause temporal changes in population connectivity and cause variations in gene flow over time (Zaccara *et al.*, 2014; Ford *et al.*, 2015; Kavembe *et al.*, 2016).

Lake Natron and Magadi were once connected and formed a larger and deeper (~ 55 m) freshwater lake, approximately 12 ka. This was intermittently connected to other rift valley lakes as part of a perennial river network (Dommain *et al.*, 2022). The lake split to form Natron and Magadi approximately 7-8 ka during the Younger Dryas aridity event, with soda conditions arising soon after as the lakes contracted (Roberts, 1993; Owen *et al.*, 2018; Dommain *et al.*, 2022). Fossil remains resembling *Alcolapia* have been found in multiple locations in the ancient lakebed sediments surrounding Lake Magadi and date from around 9-12ka (Coe, 1966; Trewavas, 1983; Roberts *et al.*, 1993; Tichy and Seegers, 1999; Dommain *et al.*, 2022). These fossils have a much larger body size than present-day *Alcolapia*, suggesting the species may have rapidly evolved a smaller body size in response to the extreme conditions. Based on nuclear and mitochondrial sequences, the Magadi species *A. grahami* is estimated to have diverged from the Natron *Alcolapia* species 7 ka – 1.55 Ma (Ford *et al.*, 2019). Given the timing of the lake split, approximately 7-8 ka, it is possible that the divergence of *A. grahami* from the other species coincided with the lake split.

While many of the factors underpinning *Alcolapia* speciation have been explored, including mate choice (Chapter 2), ecology (Ford *et al.*, 2016; Kavembe *et al.*, 2016; Chapter 3), morphology (Ford *et al.*, 2016; Kavembe *et al.*, 2016) and genetic relationships (Zaccara *et al.*, 2014; Ford *et al.*, 2015, 2019; Kavembe *et al.*, 2016), a number of questions remain regarding the timing of their origin and demographic history since the onset of divergence. Firstly, it is unclear whether the Natron radiation occurred before or after the Natron-Magadi split. A second unresolved problem is the *Alcolapia* species tree; *A. alcalica* and *A. grahami* share similar head and mouth morphology (Ford *et al.,* 2016), and a nuclear DNA dataset based on seven loci (Ford *et al.,* 2019) grouped *A. alcalica* and *A. grahami* as sister species. However, phylogenetic trees based on mtDNA (control region) (Seegers *et al.*, 2001) and SNP data (Ford *et al.,* 2015) have placed *A. grahami* as the sister species to the rest of the Natron radiation. High levels of gene flow between Natron species may obscure the species tree and cause the allopatric *A. grahami* to group as the sister species to the rest of the radiation. Furthermore, the estimation of demographic history has only been investigated between Magadi populations (Kavembe *et al.*, 2016) and not throughout the whole species flock.

In this study, we sought to model patterns of gene flow and divergence between *Alcolapia* populations by harnessing the information contained in a published restriction-site associated DNA sequence (RADseq) dataset (Ford *et al.,* 2015) and a whole genome re-sequencing dataset (WG) (Figure 4.1). Whole genome datasets can offer more robust species comparisons than is possible with RADseq and ddRADseq datasets, as more of the genome is represented. However, the RADseq dataset here provides a more comprehensive overview of *Alcolapia* diversity, with many more samples (88) and sites (11) compared to the whole genome dataset (14 samples from 6 sites). The RADseq dataset, therefore, offers opportunities for investigating relationships across multiple populations. In addition, having multiple individuals per population (4 or more) allows the inference of population relationships rather than species differences. In this chapter, we used the relative strengths of both datasets in a complementary manner to explore the same hypotheses.

Our first aim of the research presented in this chapter was to elaborate on existing findings on the species relationships, population structure, and extent of gene flow in

Alcolapia previously found using RADseq datasets (Ford *et al.*, 2015; Kavembe *et al.*, 2016) by extending the analysis to whole genome data and using additional approaches. Our second aim was to make inferences on the demographic history and patterns of gene flow in the *Alcolapia* radiation, to investigate the dynamics of their speciation over time. The results offer potential insights into how the *Alcolapia* radiation may have arisen including the role of gene flow in the radiation from its initiation to the present day.

4.2 Methods

4.2.1 Sampling and sequencing

The 14 whole genome (WG) samples (this study) and the 88 RADseq samples from Ford *et al.* (2015) were all derived from *Alcolapia* specimens collected in 2012. WG sample libraries were previously produced using TruSeq DNA Library Prep Kit and sequenced using Illumina HiSeq 2500 (A.G.P Ford, *unpublished*). The library and sequencing steps for the RADseq dataset are described in Ford *et al.,* (2015).

In addition, Illumina genome sequences of a *Oreochromis niloticus* sample (Japanese strain WL_ON_1) were downloaded from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/SRX7899814) to provide an outgroup for *f*-branch analyses (section 4.2.6).

4.2.2 Read Processing and Assembly

WG and RADseq data were processed using the same pipeline. The base quality scores of forward and reverse fastq files were first checked using *FastQC* v.0.11.9 (Andrews, 2010). Each sample's reads were aligned to the *Oreochromis niloticus* reference genome UMD1 (NCBI Assembly GCF_001858045.2) (Conte *et al.*, 2017) using *bwa-mem* v.0.7.17 (Li, 2013), then *samtools* v1.9 was used to convert the output into sorted bam files (Li *et al.*, 2009). To reduce the effect of over-represented reads as a result of biased PCR amplification, PCR duplicates were removed using *picardtools* v.1.8.0 (http://broadinstitute.github.io/picard/) before the files were indexed using *samtools* (Li *et al.*, 2009). Genome coverage of each sample was then visualised using *BEDtools* v.2.27.1 (Quinlan and Hall, 2010) with mapping quality quantified using the *samtools* "flagstat" command. Since some of the WG and RADseq samples

had low coverage, the resulting bam files were processed using different pipelines depending on the subsequent analysis. In addition, we removed four RADseq samples with fewer than 2.5 Mb of reads as per the previous study (Ford *et al.*, 2015).

For the first pipeline to calculate *f*-branch statistics for both the WG and RADseq, *GATK* v.4.1.0 (Poplin *et al.*, 2017) and *BCFtools* (v.1.10.2) (Li, 2011) were used for SNP calling and filtering. The *GATK HaplotypeCaller* program was run with soft clipped bases removed from the analysis to produce GVCF files. Each sample's GVCF file was then combined and genotyped using the *GATK CombineGVCFs* and *GenotypeGVCFs* commands to generate a single multi-sample VCF file. The resulting files were then filtered using *BCFtools* with base reads with < 5 depth coverage removed and with genotype quality and mapping quality thresholds of > 20 . For *f*-branch and demographic models, site 17 samples were removed from the RADseq dataset as this site contains morphological hybrids with uncertain species assignment (Ford *et al.,* 2015). Likewise, Lake Nakuru *A. grahami* samples were discarded as they may have been subject to demographic changes which could affect estimates. In the WG dataset, two very low coverage (2-3X mean coverage) *A. grahami* samples were excluded for *f*-branch analysis.

4.2.3 Principal Components Analysis using PCAngsd

To investigate the genetic relationships between samples and compare the WG and RADseq dataset, principal component analyses (PCA) were performed using *PCangsd* v.0.98, which is suitable for datasets with a low depth of coverage (Korneliussen *et al.*, 2014; Meisner and Albrechtsen, 2018). The following filters were applied: minimum mapping quality: 20; minimum base quality: 20; minimum minor allele frequency threshold: 0.05; *p*-value of calling a SNP: < 2e-6. The same filters were applied to the RADseq and WG samples. Separate PCAs were run with *A. grahami* included and excluded for both data sets to observe genetic differences within the Magadi-Natron and Natron systems, respectively.

4.2.4 RADseq coancestry matrix using *fineRADstructure*

To visualize the population genetic structure in more detail, *fineRADstructure* v.0.3.2 was used, which generates a co-ancestry matrix using RADseq data (Malinsky *et al.*, 2018b). *Stacks* v. 2.53 (Catchen *et al.*, 2011; Rochette *et al.*, 2019) was used to

generate the full haplotype data input files for *fineRADstructure,* using the "RADpainter" command. The *Stacks* pipeline was executed by using the "ref_map.pl" script provided in the package, which runs both the *gstacks* and *populations* programs. The sorted RADseq bam files were used as input along with a population map file containing 23 individual populations (Figure 4.1) which were specified by individuals from unique species and sites. Minimum mapping quality was set to 20 in *gstacks*. In the *populations* program -p (the minimum number of populations a locus must be present in to process a locus) was set to 10, -r (the minimum percentage of individuals in a population required to process a locus for that population) was set to 0.5 and the "--radpainter" option was selected to generate the *RADpainter* input file. Samples with more than 90% missing data were removed before input into *RADpainter*.

4.2.5 Relative rates of gene flow using *f*-branch

Patterson's *D* (or the ABBA-BABA statistic) and other related *f* statistics (such as the *f4*-ratio) are methods for quantifying the level of introgression between closely related species and populations (Green *et al.*, 2010; Durand *et al.*, 2011). However, the results of these tests can be difficult to interpret when different taxa share branches, as gene flow can lead to several correlated elevated *f4* statistics. To address this issue, Malinsky *et al.* (2018a) developed the *f*-branch ($f_b(C)$) statistic, which, given a specified tree, calculates introgression between a species *C* and branch *b* relative to the sister branch of *b.*

Dsuite (Malinsky *et al.*, 2021) was used to calculate $f_b(C)$ statistics for the WG and RADseq datasets. The calculation of the *f*-branch statistic requires a species tree. While there is strong support for *A. ndalalani* and *A. latilabris* as sister species (100% bootstrap support), the placement of *A. alcalica* is less certain (Ford *et al.*, 2015). The same phylogeny provided weak support for *A. alcalica* from the northern populations as sister to all species from the sympatric southern populations (Ford *et al.*, 2015). Likewise, Ford *et al.* (2015) found maximum support (100% bootstrap support) for *A. grahami* grouping as a sister to all other *Alcolapia* species (Ford *et al.,* 2015). However, this topology may result from extensive gene flow between sympatric species. A phylogeny based on six concatenated nuclear markers found strong support (99% bootstrap support) for the placement of *A. grahami* and the northern *A. alcalica* as sister lineages (Ford *et al.,* 2019). Therefore, this presumed species tree was used

(which is tested in section 4.2.6) with *A. grahami* and *A. alcalica* forming a clade, which is sister to the clade comprising *A. latilabris* and *A. ndalalani.* In addition, the northern *A. alcalica* (allopatric) and southern *A. alcalica* (sympatric) lineages were included as separate sister populations to measure the relative levels of gene flow between them and other taxa.

For both the WG and RADseq datasets, samples were grouped by their respective species or clade (northern and southern *A. alcalica* clades) (Figure 4.1). Using *Dsuite*, the vcf file was first split into vcf files for individual chromosome using *tabix* v.0.2.6 (Li, 2011). To calculate the *f4*-ratio, the *Dsuite Dtrios* command was used with a Jackknife window size of 50 SNPs. The resulting chromosome files were merged using the *DtriosCombine* command and input to the *Fbranch* command, which calculated $f_b(C)$ scores along with their associated *p*-values. *p*-values were corrected using the Bonferroni method to account for multiple comparisons.

4.2.6 Demographic modelling using *fastsimcoal2*

To explore and compare different evolutionary scenarios, including the history of gene flow, the relative timing of population splits and population sizes, demographic modelling was implemented using *fastsimcoal2* v.2.7 (Excoffier *et al.*, 2013, 2021). *fastsimcoal2* can model complex demographic scenarios with gene flow between multiple populations. As *fastsimcoal2* uses the Site Frequency Spectrum (SFS) for estimations, the RADseq dataset is well-suited for this analysis due to the presence of multiple samples per species.

Using this approach, the primary objectives were to:

1. Test if *A. alcalica* is nested with the other Natron species or if it is a sister species to *A. grahami*.
2. Estimate rates of gene flow between lineages and test if gene flow occurred between ancestral lineages.
3. Estimate the timing of the *A. grahami* split relative to the Natron-Magadi split.
4. Estimate the relative timing of the *Alcolapia* most recent common ancestor compared to the formation and splitting of Palaeolake Orolonga.

To address these objectives, scenarios were created with three demes to reduce computational complexity. To focus on the timing of the allopatric split, *A. grahami*, *A. alcalica* and *A. latilabris* populations were selected. *A. latilabris* and *A. ndalalani* essentially form one genetic cluster (Ford *et al.,* 2015); hence *A. ndalalani* was

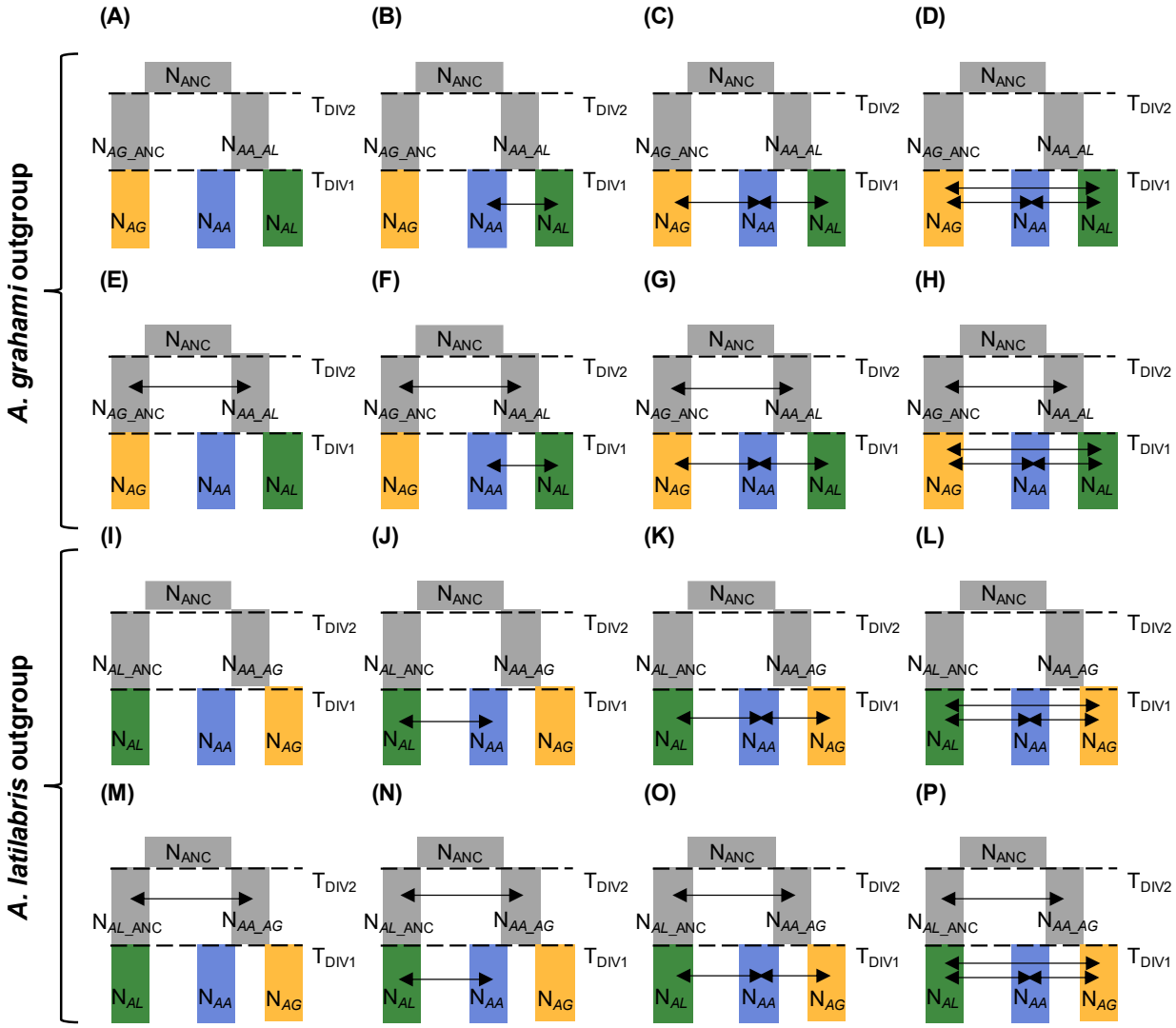


Figure 4.2. Schematic diagrams showing the demographic scenarios tested using *fastsimcoal2* with alternate topologies and patterns of gene flow. Scenarios **A-H** use a topology with Natron species as sister, whereas **I-P** use *A. grahami* and *A. alcalica* as sister species. All scenarios included coalescent parameters (T_{DIV1} and T_{DIV2}) and effective population size (*Ne*) parameters (e.g. N_{AL}, N_{AA_AG}) that were allowed to vary after T_{DIV1}/T_{DIV2} . Gene flow parameters between demes are denoted by arrows and were continuous and bidirectional. N_{AL} = *A. latilabris,* N_{AA} = *A. alcalica*, N_{AG}= *A. grahami,* N_{AL_ANC}= *A. latilabris* ancestor, N_{AA_AG}= *A. alcalica*/*A. grahami* ancestor, N_{AA_AL}= *A. alcalica*/*A. latilabris* ancestor, N_{ANC}= *Alcolapia* common ancestor.

excluded to make the analysis computationally tractable. *A. latilabris* was included rather than *A. ndalalani* due to higher sample size (n = 14 vs. n = 10) and its wider distribution in Natron (Figure 4.1).

Three-dimensional (3D) site frequency spectra were created using *ANGSD* (v. 0.933) (Korneliussen *et al.*, 2014b). First, SAF files were generated separately for each species. In *ANGSD*, base assignment qualities were computed (option '-baq 1') (Li, 2011c) and mapping quality was adjusted for excessive mismatches ('-C 50'). Only paired reads were kept ('-only_proper_pairs 1') and these were filtered for base quality values of > 20 and read mapping quality > 20. For each species, SNPs were discarded if they were present in < 80% of individuals. To remove low depth calls and potentially paralogous regions, sites were removed for each species if the combined sequencing depth of the samples was less than two times or more than 60 times the total number of samples. The *ANGSD* subprogram *realSFS* was then used to create an unfolded 3D SFS file.

Two alternative topologies were tested with either Natron taxa as sister species or *A. grahami* and *A. alcalica* as sister species (I-P). Models had two coalescent events (T_{DIV1} and T_{DIV2}) moving backwards through time. For each of these topologies, eight different gene flow scenarios were tested:

1. No gene flow.
2. Natron species gene flow after T_{DIV1}.
3. Natron and *A. alcalica-A. grahami* gene flow after T_{DIV1}.
4. Gene flow between all demes after T_{DIV1}.
5. Gene flow before T_{DIV1}.
6. Gene flow before T_{DIV1} and Natron species gene flow after T_{DIV1}.
7. Gene flow before T_{DIV1}, Natron and *A. alcalica-A. grahami* gene flow after T_{DIV1}.
8. Gene flow between all demes before and after T_{DIV1}.

These models are shown in Figure 4.2 with both Natron species as sister (A-H), and *A. grahami* and *A. alcalica* as sister species (I-P). In all cases, gene flow was modelled as continuous and bidirectional. Prior search ranges for the best-fitting scenario for the *fastsimcoal2* analysis are provided in Table A4.2.

For each of these scenarios, *fastsimcoal2* was run with 50 ECM cycles and with 100,000 simulations, a total of 50 times. The option '-foldedSFS' was used to fold the unfolded SFS produced by *realSFS*. Entries with less than 10 SNPs were pooled together ('--minSFSCount 10') to avoid overfitting. As estimates of the *Alcolapia* per

generation mutation rate per site (μ) are unavailable, we used the point estimate of μ = $3.5x10^{-9}$ derived from Malawi cichlids (Malinsky *et al.,* 2015).

The run which had the highest estimated likelihood was extracted for each scenario. These runs with the best fit were then compared between competing scenarios using the Akaike Information Criterion (AIC) (Akaike, 1974). AIC values may be biased due to the presence of linkage between sites (Excoffier *et al.*, 2013); therefore for each scenario, a likelihood distribution was generated from 100 estimations of the coalescent simulation and with 1 million coalescent iterations. Both ΔAIC (the difference in AIC with the best-fitting model) and overlaps in likelihood distributions were used to test for significant differences between the best-fitting models from each competing scenario.

For the best-fitting scenarios, confidence intervals were obtained using a nonparametric bootstrap approach. Using the *realSFS* subprogram, 100 SFS files were created with the "-bootstrap" option. *fastsimcoal2* was run on these bootstrapped SFS files using the same template and estimation files from the best-fitting scenario with identical options. After obtaining the best-fitting model for each of these bootstrapped SFS, 95% confidence intervals were calculated using the *R* package *boot* v.1.3.20 (Canty and Ripley, 2022).

4.3 Results

4.3.1 WG re-sequencing and mapping to the *O. niloticus* reference genome

After the alignment and removal of PCR duplicates, 953 million reads (97.8 – 99.3% of reads) of the WG dataset were mapped to the *O. niloticus* genome. This compares to a mean mapping rate of 89% in the RADseq dataset which was previously found by Ford *et al.* (2015). The coverage of the WG re-sequenced datasets was highly variable, with a range of 3.3x to 37.5x mean coverage in the lowest and highest coverage samples, respectively (Table A4.1).

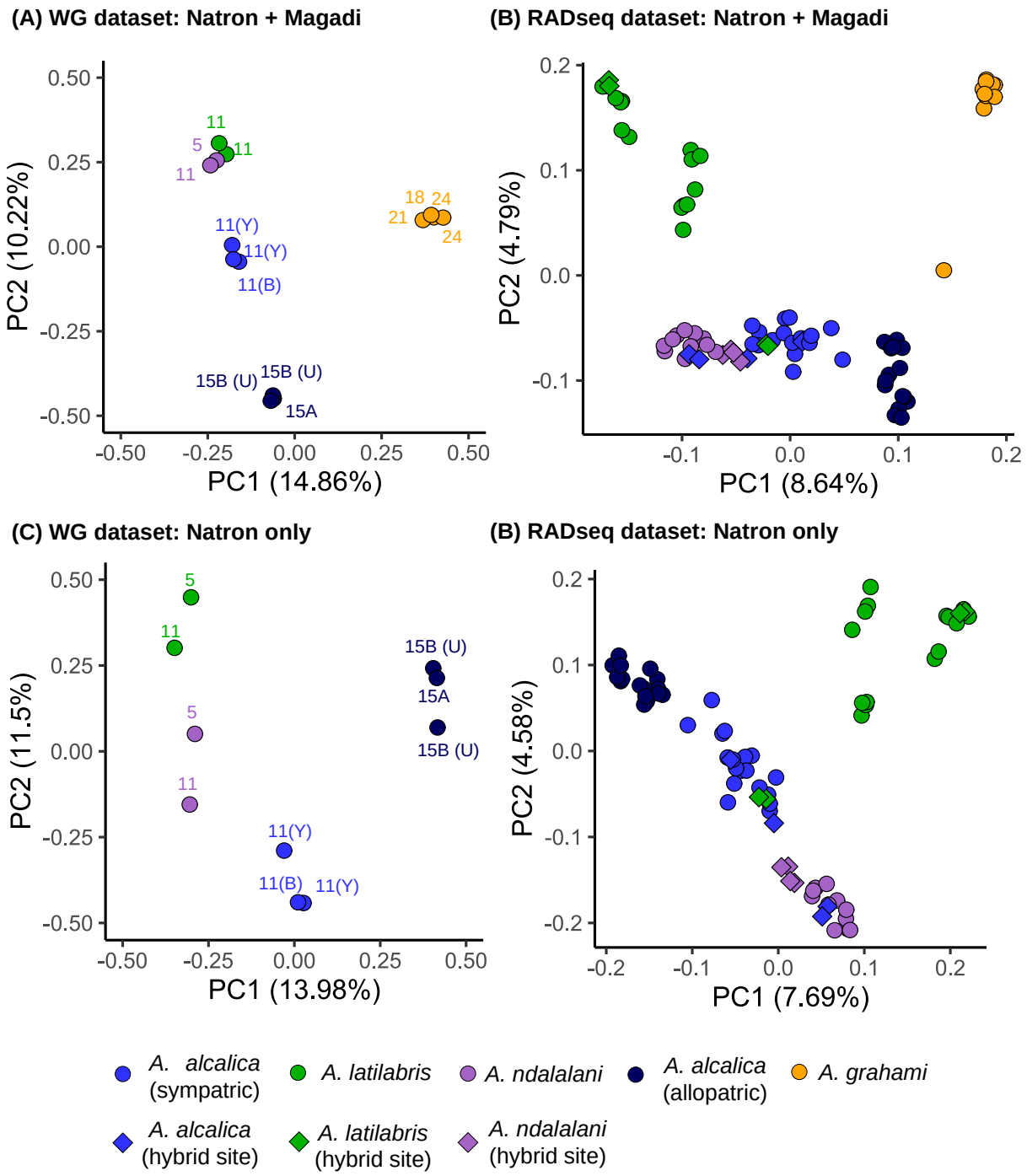


Figure 4.3. Principal Component Analysis (PCA) of shows similar species relationships in *Alcolapia* when using WG (whole genome) and RADseq datasets. Generated using *PCAngsd.* (**A**) WG PCA of Natron and Magadi *Alcolapia*, (**B**) RADseq data of Natron and Magadi *Alcolapia*, (**C**) WG Natron only, (**D**) RADseq Natron only. Colours denote the species/clade with diamond-shaped points corresponding to site 17 individuals that contain morphological hybrids (Ford *et al.,* 2015). For WG samples, numbers refer to the collection site with morphotypes in brackets: U = upturned mouth morph, B = blue morph, Y = yellow morph.

4.3.2 Species relationships

We found similar genetic clustering between *Alcolapia* species using both datasets (Figure 4.3A, B). PC1 and PC2 explained 25.1% and 13.4% of the variation in the WG and RADseq datasets, respectively. *Alcolapia grahami* was well separated from the other species in each dataset, corresponding to previous findings that *Alcolapia* species cluster by lake (Ford *et al.*, 2016). Moreover, *A. latilabris* was also the most distinct from the other populations in the RADseq dataset, but to a lesser degree in the WG datasets (Figure 4.3A, B).

We found sister species *A. latilabris* and *A. ndalalani* to cluster more closely compared to the other species in the WG dataset (Figure 4.3A, C). Likewise, site 15A and 15B *A. alcalica* samples, which were previously found by Ford *et al.* (2015) to cluster separately from the southern sympatric *A. alcalica* samples, were also well separated in the WG dataset (Figure 4.3C, D). In the separate PCA of just the Natron samples, some evidence of intraspecific variation by site can be observed between WG samples, with sites 5 and 11 varying along the PC2 axis (Figure 4.3C). We found no evidence of genetic clustering between the same intraspecific morphs using PCA; both the upturned and terminal mouth *A. alcalica* morphs from sites 15A and 15B clustered together, as did the blue and yellow *A. alcalica* morphs from site 11 (Figure 4.3A, C).

4.3.3 Population structure by site

After filtering with stacks, 518,917 variant sites remained for *fineRADstructure* analysis, and eight RADseq samples which had over 90% missing data were dropped. Our findings elaborate on existing results by Ford *et al.* (2015) as we observed clustering by species with some population structuring by site (Figure 4.4). The co-ancestry matrix revealed that *A. grahami* had high levels of within-species relatedness and were the most distinct from the rest of the Natron species. Pronounced site-specific population clustering was most apparent in all species from site 5, northern *A. alcalica* from sites 6 and 15, and in all *A. latilabris* populations except for hybrid site 17. Other than at hybrid site 17 and site 5, both sympatric *A. alcalica* and *A. ndalalani* showed less distinct site-specific relatedness in comparison to *A. latilabris*. Other than

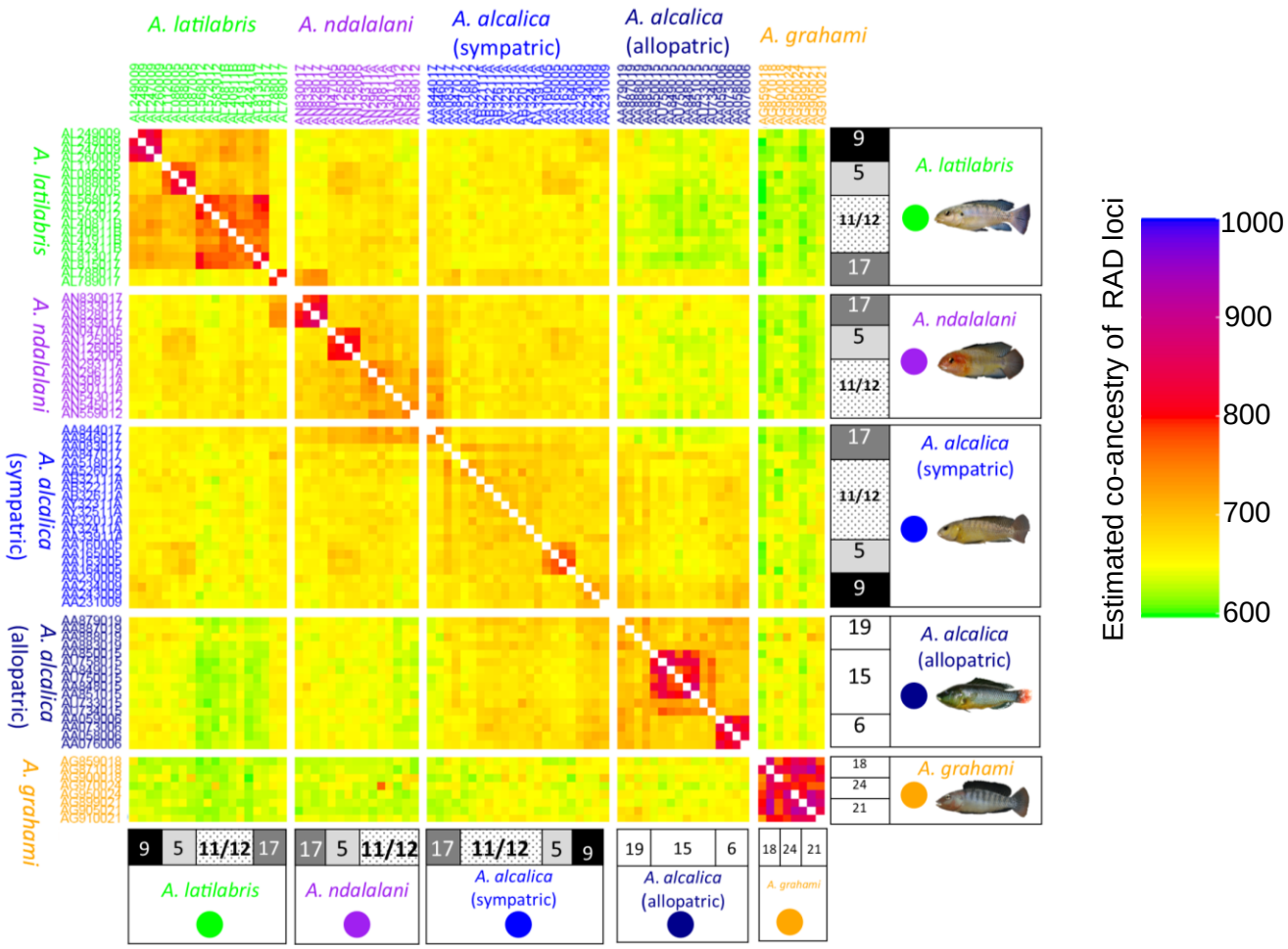


Figure 4.4. Co-ancestry matrix generated from *fineRADstructure* highlights the elevated rates of gene flow between Natron *Alcolapia* species and populations. Individual sample ID is listed at the top of the columns and to the left of each row. Blocks are divided by species/clade with numbers on the right and bottom of the matrix corresponding to the site the individual was collected from. Colour corresponds to degree of genetic relatedness with red to blue indicating high levels of relatedness and yellow to green low levels of relatedness. Haplotype similarity between individuals is based on 518,917 SNPs in the RADseq dataset.

site 5 individuals, sympatric *A. alcalica* showed less site-specific co-ancestry, suggesting these populations may be less localised.

Results for site 17 suggest that some samples cluster to different species than to those which they were assigned. There is a unique cluster between all (hybrid) site 17 *A. ndalalani* individuals and two out of the four site 17 *A. latilabris* individuals. Contrastingly, the other two site 17 *A. latilabris* individuals cluster with site 11B/12 *A.*

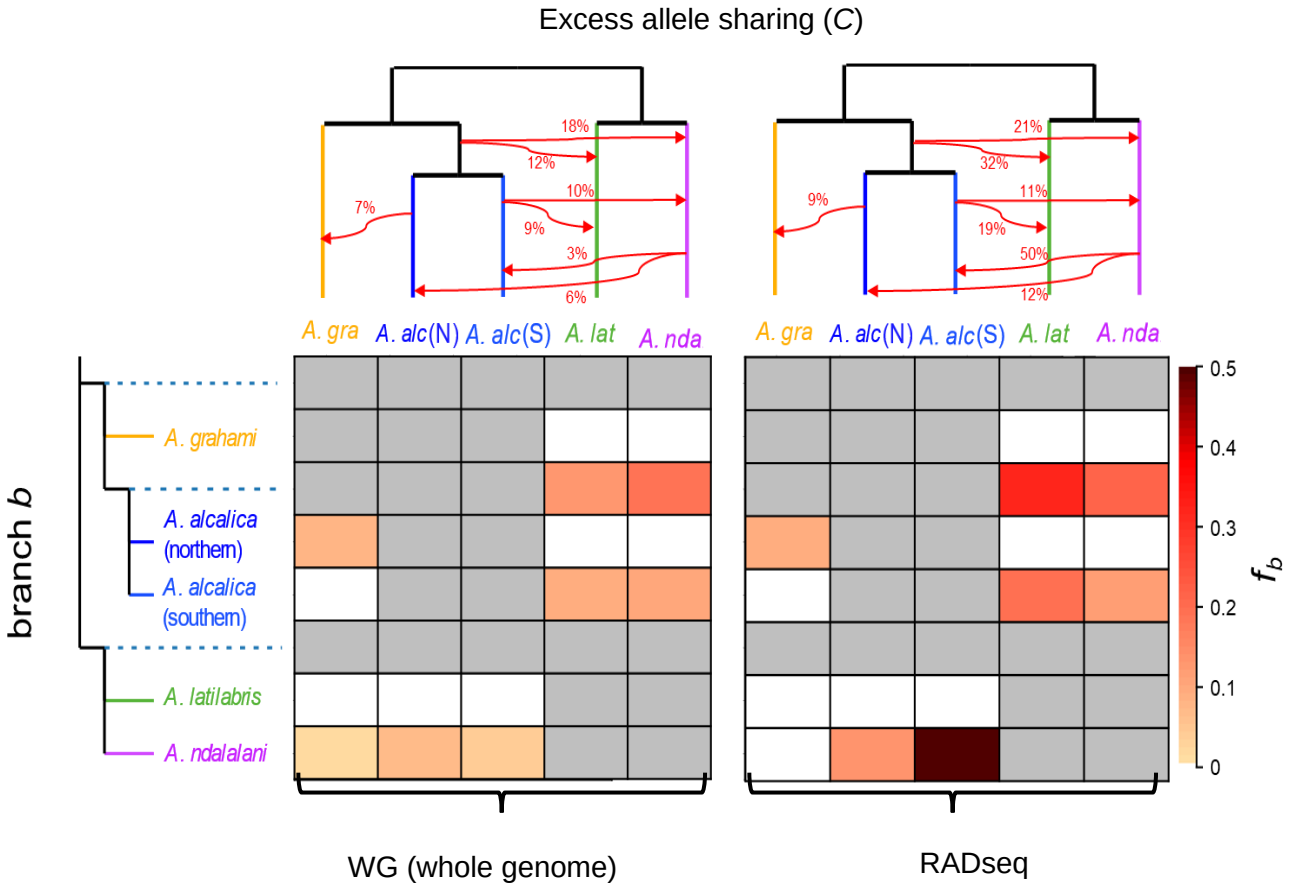


Figure 4.5. *f*-branch ($f_b(C)$) statistics highlight the excess allele sharing between populations and branches using both (**A**) whole genome and (**B**) RADseq datasets. The value of the $f_b(C)$) statistic is given by the colour of the cell in the matrix which compares the proportion of excess sharing of derived alleles with different species/populations *C* (*x* axis) relative to its sister branches (*y* axis). The red arrows correspond to the proportion of admixture between a given tip/branch relative to its sister tip/branch. Only significant results (*p*-value < 0.05 after Bonferroni correction) and > 1% excess allele sharing are given. Grey areas denote pairwise comparisons that are not possible on account of the tree topology.

latilabris. A similar pattern is visible in site 17 *A. alcalica*, where two individuals show higher relatedness to *A. ndalalani* while the other two are closer to *A. alcalica* (Figure 4.4).

Elevated interspecific relatedness is most evident at site 5, suggesting increased gene flow at this site. Likewise, site 9 *A. alcalica* and *A. latilabris* have higher relatedness with individuals at sites 11 and 12 compared to site 5, despite these populations being situated closer around the lake perimeter (Figure 4.1).

4.3.4 Relative gene flow rates with *f*-branch

Using the presumed species tree, both the WG and RADseq datasets showed similar patterns of excess allele sharing between taxa, but levels were generally more extreme in the RADseq dataset (Figure 4.5). As predicted, there was evidence of increased gene flow in sympatry, with excess allele sharing recorded with other Natron species between the *A. alcalica* branch relative to the *A. grahami* branch. Moreover, there was evidence that introgression rates are higher among the sympatric sub-populations: southern *A. alcalica* from sympatric sites shared more alleles with Natron species, whereas allopatric *A. alcalica* had excess allele sharing with *A. grahami* (Figure 4.5).

Among the sympatric species, *A. ndalalani* had excess allele sharing with both *A. alcalica* lineages relative to *A. latilabris*, but the extent of this varied between datasets. For instance, in the RADseq dataset we found very high excess allele sharing between *A. ndalalani* and sympatric *A. alcalica* (50%); in the WG dataset this was significantly lower (3%) (Figure 4.5). We note that the reported level of introgression in the RADseq dataset should be treated with caution, as performance is dependent on the number of SNPs, and the authors of the package found that RADseq estimates were only accurate in < 40% of simulations (Malinsky *et al.*, 2021).

4.3.5 Demographic inference

Of all the demographic scenarios, the best fitting model supported *A. latilabris* as the sister group deme to *A. grahami* and *A. alcalica*. This had the best fit both in terms of AIC (2,047,666) and in $\log_{10}$Likelihood (-196,775) (Figure 4.6; Table A4.3). In the absence of gene flow, a topology with *A. latilabris* as the sister group to the other taxa had a worse fit (ΔAIC = 4821) compared to one with *A. grahami* as the sister group to other taxa (ΔAIC = 2312) (Figure 4.6). However, the inclusion of gene flow between *A. alcalica* and *A. latilabris* consistently improved the fit of this topology. The best-fitting scenario included gene flow between Natron species (*A. latilabris* and *A. alcalica*), as well as between *A. alcalica* and *A. grahami* after the most recent species split (Figure 4.6).

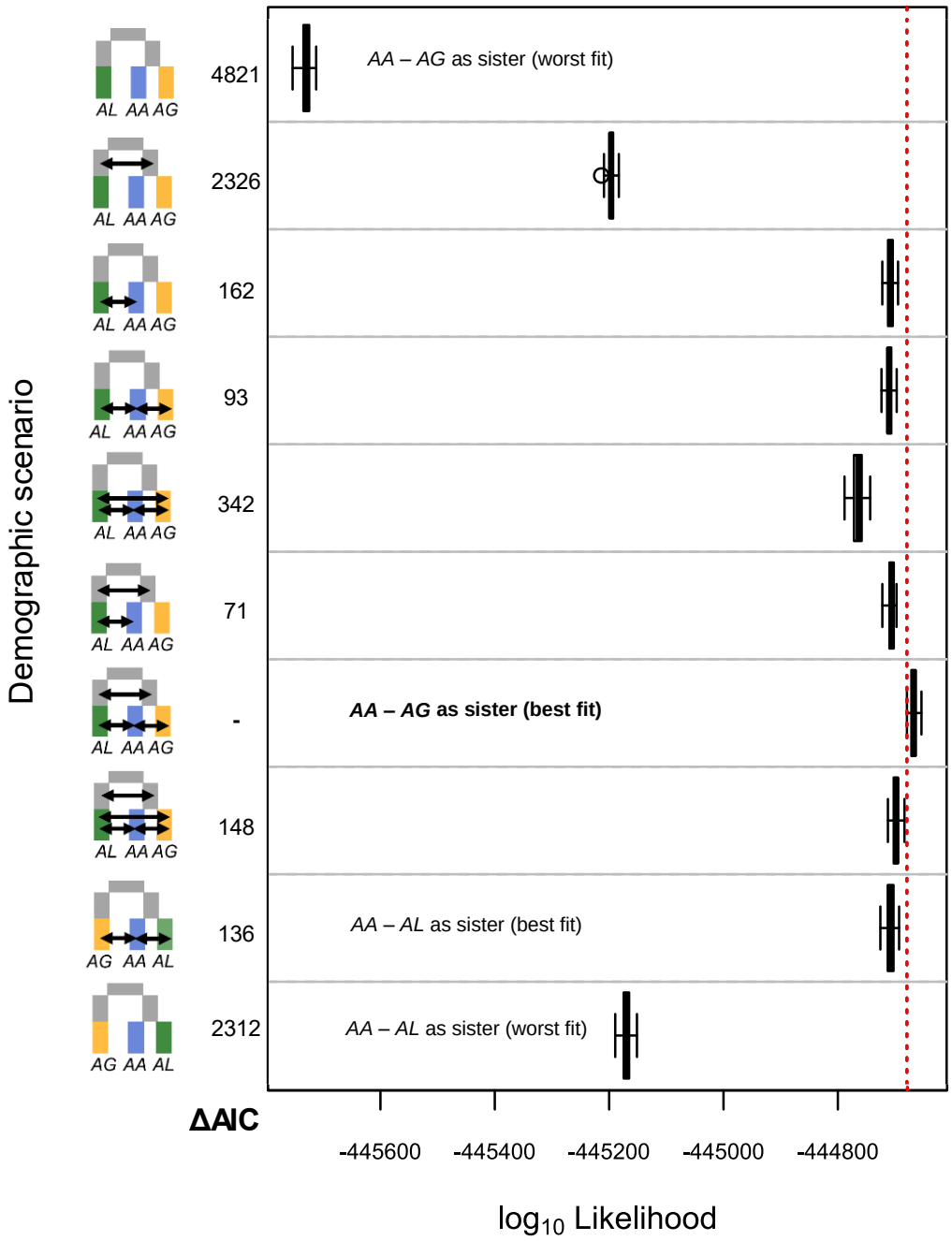


Figure 4.6. Comparison of demographic scenarios supports the paraphyly of Lake Natron *Alcolapia* species and gene flow. Demographic scenarios modelled using coalescent simulations (*fastsimcoal2)* supported a topology of *A. alcalica* and *A. grahami* (Lake Magadi) as sister species to *A. latilabris* (both by AIC and $\log_{10}$Likelihood). Shown are all gene flow scenarios using the *A. alcalica-A. grahami* sister species topology along with the best and worst fitting scenarios of the Natron sisters species' topology (on the bottom two rows). The best-fitting scenario (in bold) included gene flow between all lineages except for *A. grahami* and *A. latilabris*. *AL = A. latilabris, AA = A. alcalica*, *AG = A. latilabris.* The red dotted line is the lowest simulated $\log_{10}$Likelihood value for the best-fitting model which did not overlap with the distributions of other scenarios.

In the most likely model, the timing of the divergence of *A. grahami* and *A. alcalica* (T_{DIV1}) was estimated at 11,123 generations ago (95% Confidence Interval (CI): 10,166-12,500 generations), whereas divergence of the ancestral deme was estimated to have occurred 331,458 generations ago (95% CI: 272,491-447,020) (Figure 4.7). Assuming a generation time of six months as reported in other *Oreochromis* species (Philippart and Ruwet, 1982), the estimate for the *A. grahami-A. alcalica* split occurred 5,589 (95% CI: 5538-6273) years ago, whereas the ancestral species split occurred 185,719 (95% CI: 136,246-222,895) years ago.

Corresponding with the current distributions of the species, *A. alcalica* had the largest estimated population size (*Ne* = 79,001, 95% CI: 74,597-85,996), followed by *A. grahami* (*Ne* = 44,857, 95% CI: 42,562-47,891) and *A. latilabris* (*Ne* = 24,985, 95%, CI: 23,066-27,839; Figure 4.7). Before T_{DIV1}, the *A. latilabris* ancestor deme was inferred to have a larger combined population size compared to after the split (N_{AL_ANC} = 144,424, 95% CI: 120,572-170,179), whereas the *A. alcalica - A. grahami* ancestral deme had a similar combined size (N_{AA_AG} = 180,092, 95% CI: 152,366-216,919), but the population of the common ancestor was reduced (N_{ANC} = 65,884, 95% CI: 44,015-123,256; Figure 4.7).

Significant levels of admixture were estimated between Natron species (Figure 4.7). *A. alcalica* received a higher level of migration with respect to relative population sizes, with a probability of an allele migrating of 1.78×10^{-4} (95% CI: $1.47\text{-}2.77 \times 10^{-4}$), which corresponds to 4.45 migrants per generation (95% CI: 3.59-6.92). *A. latilabris* received an allele migration probability of 4.51×10^{-5} (95% CI: $3.22\text{-}9.53 \times 10^{-5}$), equivalent to 3.56 migrants per generation (95% CI: 2.55-7.53). Gene flow estimates between *A. alcalica* and *A. grahami* were largely unidirectional, with *A. alcalica* receiving an allele migration probability of 1.56×10^{-5} (95% CI: $1.16\text{-}2.17 \times 10^{-5}$) compared to 2.76×10^{-6} in *A. grahami* (95% CI: 1.04×10^{-7}-1.60×10^{-5}; Figure 4.7). The best-fitting scenario did not include migration parameters between *A. latilabris* and *A. grahami* (Figure 4.6). Additionally, point estimates for inferred levels of gene flow were consistently low across models that included these parameters (range of 2.9×10^{-6} to 5.5×10^{-9}), indicating either weak or absent gene flow between these species (Table A4.3).

Inferred levels of gene flow were significantly higher between the ancestral lineages, with migration probabilities of 6.27×10^{-4} (95% CI: 3.37×10^{-4}-1.24×10^{-3}) from N_{AL_ANC}

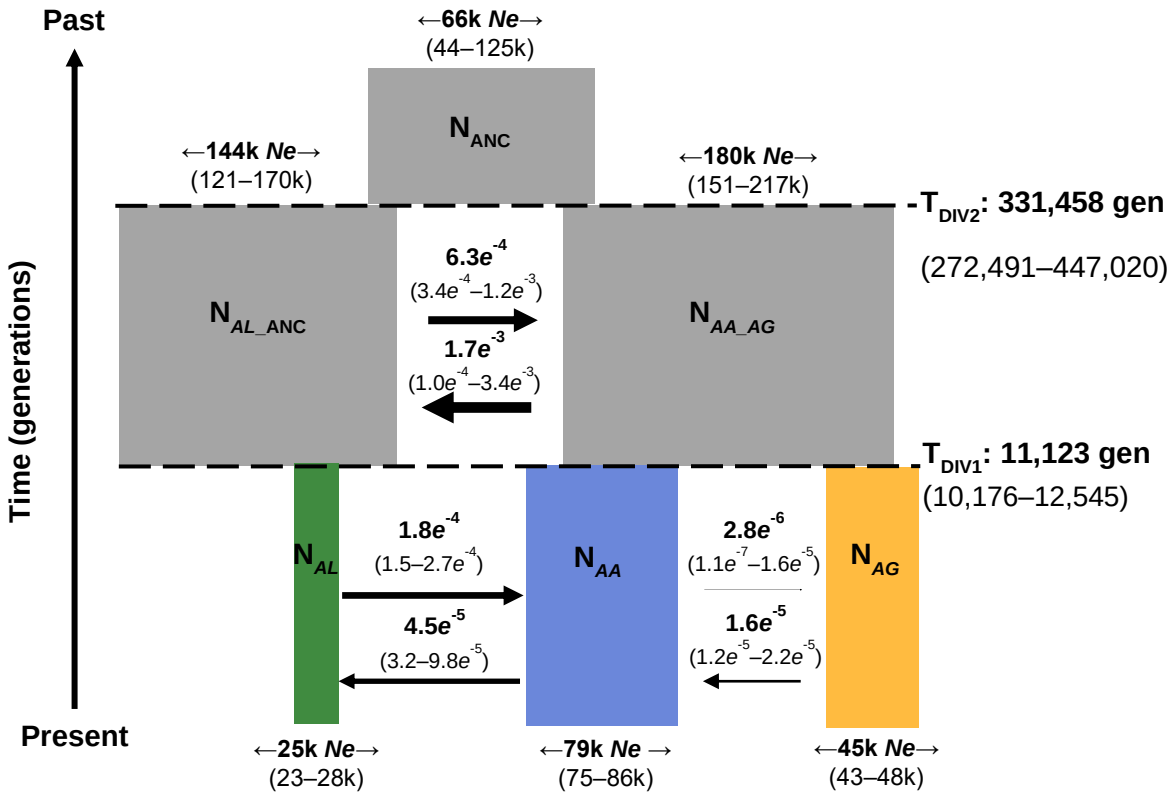


Figure 4.7. Schematic figure showing divergence times, population sizes and gene flow in the most supported demographic model generated in *fastsimcoal2*. N_{AL} = *A. latilabris,* N_{AA} = *A. alcalica*, N_{AG}= *A. grahami,* N_{AL_ANC} = *A. latilabris* ancestral deme, N_{AA_AG} = *A. alcalica* and *A. grahami* ancestral deme. Values above the arrows denote inferred migration rates and refer to the probability of an allele migrating per generation. Values in bold are the mean bootstrap parameter estimates whilst the 95% confidence limits are provided in brackets.

(*A. latilabris* ancestral deme) to N_{AA_AG} (*A. alcalica* and *A. grahami* ancestral deme) and 1.7 x 10^{-3} (95% CI: 1.02 x 10^{-4}-3.31 x 10^{-3}) from N_{AA_AG} to N_{AL_ANC} (Figure 4.7). However, high admixture rates were not present across all other models that included ancestral gene flow, such as the second-best supported model by AIC (ΔAIC =71), which had migration probabilities of 3.5 x 10^{-8} and 5.7 x 10^{-5} (Figure 4.6; Table A4.3).

4.4 Discussion

4.4.1 Main findings

Hybridisation is increasingly recognised as a driving force in adaptive radiations (Berner and Salzburger, 2015; Jamie and Meier, 2020; Marques *et al.,* 2020). Analyses of genomic datasets have revealed gene flow to play an important role during

the early stages of speciation in several young radiations, such as in Darwin's finches (Almén *et al.,* 2016), in three-spined sticklebacks (Bassam *et al.,* 2018), and in *Mimulus* monkeyflowers (Stankowski and Striesfield, 2015). Speciation with gene flow is particularly common in cichlid adaptive radiations (Brawand *et al.,* 2015; Malinsky *et al.,* 2018a; Irisarri *et al.,* 2018) and is especially prevalent between *Alcolapia* species from Lake Natron (Ford *et al.,* 2015). Here, we found further evidence of gene flow between sympatric Natron species using both whole genome and RADseq datasets, providing additional insights into the relative rates of gene flow between species and sub-populations. These findings suggest varying interconnectedness between Natron spring sites and that populations are not entirely panmictic. Using demographic modelling, we provided support for the paraphyly of Natron species, with the single Magadi species *A. grahami* grouping as sister to *A. alcalica* and splitting after the Natron-Magadi lake split ~ 8 ka. Our estimates also suggested a much earlier origin of the *Alcolapia* species flock, with the common ancestor occurring ~ 330 (272-447) thousand generations ago, which predates the most recent formation of a joined Magadi-Natron lake basin, 12.4 ka (Figure 4.1). High levels of inferred historical gene flow between ancestral populations demonstrate that hybridisation has been an important feature in the *Alcolapia* radiation.

4.4.2 Population structure and gene flow between Natron populations

Consistent with previous findings, we found that *Alcolapia* species formed distinct genetic clusters both in the WG samples and RADseq dataset (Ford *et al.*, 2015, 2016). Likewise, there was also evidence of further within-species genetic clustering by geography. The lack of clustering between distinct intraspecific *A. alcalica* colour and mouth morphs further supports a fine-scale genetic basis for these phenotypes (Maan and Sefc, 2013; Ford *et al.*, 2015). The RADseq results also elaborate on previous work showing that population structure varies between species and between sub-populations of species (Ford *et al.*, 2015; Kavembe *et al.*, 2016). For instance, we found that *A. latilabris* populations tended to have much higher co-ancestry with individuals of the same populations relative to *A. alcalica*, potentially resulting from lower dispersal rates in this species (Seegers *et al.*, 2001). Likewise, locations such as site 5 showed evidence of elevated interspecific gene flow relative to other sympatric sites, suggesting that migration and admixture rates may vary across the

radiation. While sub-populations are speculated to experience intermittent periods of connectivity during seasonal flooding (Seegers and Tichy, 1999; Zaccara *et al.*, 2014; Kavembe *et al.*, 2016), the degree of relatedness between populations is not necessarily predicted by their proximity (Ford *et al.*, 2015). Our findings echoed this, with site 9 individuals having higher relatedness to site 11 and 12 populations than site 5 individuals despite their greater distance.

Intra- and inter-population genetic differentiation appeared lower in *A. grahami* relative to the other samples, although we only had samples from three populations. While Kavembe *et al.* (2016) found population structuring between separate lagoons and the satellite lake Little Magadi (Figure 4.1), the Magadi populations here originate from a joint lagoon and group together in the same clade. Similarly, the introduced *A. grahami* individuals from Lake Nakuru do not cluster separately from the Magadi samples. It is unknown from which site in Magadi the first colonising individuals were sourced from, but as there have only been 60-70 years of separation, significant divergence may be negligible (Vareschi and Vareschi, 1979).

Natron *Alcolapia* species share the same habitats and lack the ecological depth stratification typical to many other young cichlid radiations (Recknagel *et al.*, 2014; Malinsky *et al.*, 2015). However, species still vary in their distribution (Figure 4.1), and there is also anecdotal evidence of some spatial segregation of species within sites (Seegers *et al.*, 2001; G.F. Turner *personal communication*, 16th July 2022). Alongside these observations, our findings of population genetic structuring by site present a relatively complex picture of the dynamics of gene flow among Natron *Alcolapia* species. A lack of complete panmixia indicates that they do not follow strict definitions of sympatry (Fitzpatrick *et al.*, 2008; Bird *et al.*, 2012). Instead, they may best be described as “mosaic sympatric”, where species coexist in intertwining patches (Mallet *et al.*, 2009). Within their distribution, the degree of co-occurrence may vary from random assortment in some patches to spatial structuring in others. A similar pattern can be found in Nicaraguan crater lake cichlids and may be a common form of speciation when geographic barriers are absent (Barluenga *et al.*, 2006; Mallet *et al.*, 2009).

4.4.3 Demographic modelling supports the paraphyly of Natron species

The demographic scenario with the best fit supported *A. grahami* and *A. alcalica* as sister species when gene flow was included, indicating the paraphyly of lake Natron species. In the absence of migration, the scenario with *A. latilabris* as sister to *A. alcalica* was a worse fit than the alternative, highlighting the effect that gene flow can have on obscuring the species tree. While previous analyses have found inconsistent phylogenies of Natron species (Seegers *et al.*, 2001; Ford *et al.*, 2015; Astudillo-Clavijo *et al.*, 2023), strong support for allopatric *A. alcalica* and *A. grahami* as sister species was previously found using six nuclear loci (Ford *et al.*, 2019). In addition, this topology also corresponds with groupings in morphospace: *A. grahami* and *A. alcalica* share terminal mouth shapes, whereas *A. latilabris* and *A. ndalalani* have more derived downturned mouth morphology (Ford *et al.*, 2016).

Using a generation time of 6 months (Trewavas, 1983), the estimated divergence time leading to the *A. alcalica-A. grahami* and *A. latilabris-A. ndalalani* clade (165 ka, 95% CI: 136-223 ka) corresponds to before the splitting of palaeolake Orolonga to form Natron and Magadi *ca*. 8 ka (Roberts *et al.*, 1993; Dommain *et al.*, 2022). This splitting time and inferred gene flow from *A. grahami* to *A. alcalica* (5.5-6.2 ka) is inconsistent with an allopatric divergence, although the upper estimate is ~ 800 to 1800 years from the lake split and estimated gene flow was much lower than between other demes. Excess allele sharing was also recorded between allopatric *A. alcalica* samples from the more northern parts of Lake Natron that are situated closer to Lake Magadi.

In addition, after Natron and Magadi split, high water levels occurred in Magadi until at least 7.7 ka, reaching their current levels around 5.5 ka. During this early period, Magadi was even connected to Palaeolake Siriata, situated approximately 11 km from Magadi (Dommain *et al.*, 2022). Given that Natron is situated at the downstream extent of the Holocene “Southern River” (Dommain *et al.*, 2022), it is plausible that Natron received some downstream migration from Magadi populations during wet periods after the lakes split. It should also be emphasised that there is considerable uncertainty in our divergence estimates as scaling generation time to years requires a reliable estimate of mutation rate and generation time (see section 4.4.5).

4.4.4 Divergence estimates indicate a much older origin of the *Alcolapia* radiation

Divergence estimates suggest that the *A. latilabris-A. ndalalani* and *A. alcalica-A. grahami* ancestral lineages arose from a common ancestor between 272 to 447 thousand generations ago, which scales from 136 to 224 ka, assuming a 6-month generation time (Trewavas, 1983). The lower estimates for these coincide even before the connected period of Palaeloake Orolonga and the southern river during the early Holocene from 12.4 – 8.2 ka (Dommain *et al.*, 2022), although still occurs within the timeline of the formation of the Natron-Magadi basin approximately 700 ka (Eugster, 1986). Based on these results, we can conclude that *Alcolapia* diversification predates the formation of Natron and Magadi. Significant inferred admixture between ancient lineages also suggests they were in contact for some of this period.

One possible scenario is that the *A. latilabris* and *A. alcalica-A. grahami* ancestors existed sympatrically in the freshwater palaeolake. The greater depth of this lake (~55m) means that divergence along an ecological depth gradient or between the shoreline and pelagic habitats is feasible (Seehausen and Wagner, 2014; Ford *et al.*, 2015). The downturned mouth morphology common to the *A. latilabris-A. ndalalani* clade may have also arisen during this period (Ford *et al.*, 2015), although evidence of this in the fossil record is lacking, with fossils reported from Magadi so far found to possess the terminal mouth morphology that is typical of other *Oreochromis* species (Coe, 1966; Butzer *et al.*, 1972; Hillaire-Marcel *et al.*, 1986; Tichy and Seegers, 1999; Dommain *et al.*, 2022). However, to our knowledge, intact cichlid fossils from Natron have not been reported, although some isolated remains (vertebrae and lower pharyngeal jaw) have been recovered (J.J. Day, *personal communication,* 7th February 2023). After the lakes contracted and hyperalkaline and saline conditions emerged, Magadi and northern Natron sites may have lacked the physical characteristics to maintain the stable coexistence of species (Ford *et al.*, 2016). Alternatively, based on the current distribution of species, it is possible that the *A. latilabris-A. ndalalani* ancestor was initially localised to the southern parts of the lake, with populations subsequently persisting in this region after the lake split.

An older divergence time of *Alcolapia* challenges the hypothesis of the rapid nature (< 10,000 years) of the Natron radiation, which was initially based on the estimated timing

of the lakes' geological split (Williamson *et al.,* 1993; Tichy and Seegers, 1999). Climatic and hydrological changes in the rift valley likely created repeated periods of river and lake separation and connectivity, before and after the *Alcolapia* common ancestor (Eugster, 1986; Hillaire-Marcel *et al.*, 1986; Williamson, 1993; Owen *et al.*, 2019; Dommain *et al.*, 2022). Ancestral hybridisation is thought to be a major driver of cichlid adaptive radiations, with ancient admixture generating polymorphic variation in descendants (Genner and Turner, 2012; Meier *et al.*, 2017a; Irisarri *et al.*, 2018; Poelstra *et al.*, 2018). This process of ancestral hybridisation has been recorded in other lake cichlid radiations and occurred as a result of the repeated colonisation and invasion of riverine species into lacustrine populations (Joyce *et al.,* 2011; Poelstra *et al.,* 2018).

The novel reassembly of old-standing genetic variation may be a potent driving force of rapid speciation (Marques *et al.*, 2019). Based on the phylogenetic placement of *Oreochromis* species, adaptation to extreme soda conditions likely evolved once in *Oreochromis*: in the lineage containing *Alcolapia* and their close relative *O. amphimelas*, a species endemic to several lakes situated approximately 100km to the south of Lake Natron (Trewavas, 1983; Ford *et al.*, 2019). Additionally, *Alcolapia*-like fossils have been recorded in the Ngorora Formation of the Tugen hills, which is approximately 285km to the north of Lake Magadi. These fossils most resembled *Alcolapia* as opposed to other *Oreochromis*, yet occurred in alkaline palaeolakes during the middle Miocene from 10-13 Ma (Penk *et al.,* 2019; Kevrekidis *et al.*, 2020). This raises questions about the extent to which both species and adaptive traits arose independently in Natron and Magadi, or whether they resulted from ancient standing genetic variation. Indeed, differences in adaptive traits between *A. grahami* and *A. alcalica* have been recorded, such as the extent of ureotelism, an essential adaptation for surviving in soda conditions. Research on aquarium populations found that *A. alcalica* had reduced ureotelism and differential expression of ammonia transport genes (White *et al.*, 2022). Uncovering the genetic basis of species phenotypes and other adaptive traits may shed some light on their route of origin.

4.4.5 Asymmetrical gene flow among Natron species

Using demographic models, we found evidence of gene flow for much of the duration of the *Alcolapia* radiation, mirroring findings from other recent cichlid radiations (Martin

et al., 2015; Meier *et al.*, 2017b; Malinsky *et al.*, 2018a). The migration probabilities are comparable to those found in other sympatric contexts, such as in the sympatric Midas cichlid complex (7.5×10^{-5} to 8.8×10^{-5}) (Kautt *et al.*, 2016) and between sympatric *Pundamilia* species in Lake Victoria (4.3×10^{-4} to 1.7×10^{-3}) (Meier *et al.*, 2017b). We note that these migration estimates are based on modelling continuous gene flow since divergence and do not consider fluctuations in gene flow rates over time.

Results from demographic models suggest that rates of gene flow were asymmetric between species. Asymmetric gene flow often occurs when there are differences in population sizes, with the larger source population sending more migrants into the scarcer sink population than the opossite (Wirtz, 1999; Nevado *et al.*, 2011). This form of unidirectional gene flow has been recorded between other cichlid species using demographic modelling (Meier *et al.*, 2017b; Lewanski *et al.*, 2021). However, unlike in those examples, gene flow was significantly stronger in the *A. latilabris* to *A. alcalica* direction than in reverse, despite the smaller population size of *A. latilabris*. One cause for this could be the omission of *A. ndalalani*, the sister species of *A. latilabris*. Unaccounted changes to the SFS may result if *A. alcalica* receives alleles from *A. ndalalani* that are also shared with *A. latilabris*.

There are several possible biological explanations for the asymmetrical gene flow suggested in this study. The presence of physical barriers can cause migration to be skewed in one direction, such as from strong river currents and waterfalls (Hänfling and Weetman, 2006; Fitzpatrick *et al.*, 2016). Observations of wild populations have found that *A. latilabris* are more prevalent in the upper courses of streams and in more rocky habitats (Seegers *et al.*, 2001; G.F. Turner, *personal communication*, 16th July 2022). In principle, a net emigration of *A. latilabris* towards *A. alcalica* lagoon populations could cause asymmetrical gene flow. Alterations to habitat and population distribution may also lead to unidirectional gene flow. For instance, lake level fluctuations are believed to have shaped unidirectional introgression between distinct *Tropheus moorii* colour morphs by causing local extinctions in source populations (Egger *et al.*, 2010; Sefc *et al.*, 2017). In Natron, mass die-offs have been reported during heavy seasonal flooding when trona deposits dissolve, creating anoxic and hypersaline conditions (Seegers and Tichy, 1999; Wilson *et al.*, 2004). The impacts

likely differ between lagoons and spring sites, potentially causing an imbalance in the mortality rates between species.

Unevenness in the strength of intrinsic reproductive barriers between species could also lead to unidirectional admixture (Nevado *et al.*, 2011). Asymmetries in assortative mating have been documented in other cichlids and can result from differences in the strength of female mate choice or male dominance between species (Egger *et al.*, 2010; Nevado *et al.*, 2011; Sefc *et al.*, 2015; Van Steenberge *et al.*, 2022). It is possible that *A. latilabris* may have stronger mate choice barriers compared to *A. alcalica* females, thereby causing asymmetrical admixture. While we lacked sufficient power to test for this, in Chapter 2 we found some indications that *A. latilabris* females may exhibit greater assortative mate choice compared to *A. alcalica* females in aquarium mating trials. While half (4/8) of the *A. alcalica* females mated with *A. latilabris* males, none of the *A. latilabris* females (0/9) mated with *A. alcalica* females (Chapter 2.). Post-zygotic effects that differentially affect reciprocal crosses can also lead to gene flow asymmetries (Bolnick *et al.*, 2008; Arntzen *et al.*, 2009; Nevado *et al.*, 2011), but this has not yet been studied in *Alcolapia*.

4.4.6 Scaling divergence estimate to years and limitations

We caution that there are substantial uncertainties in our divergence estimates when scaling generation times to years. Divergence time estimates are scaled by generation time, which is unknown in *Alcolapia*. Sexual maturation times in *Alcolapia* are estimated to be much shorter (42-60 days) (Coe, 1966; Wilson *et al.*, 2004) in comparison to other *Oreochromis* (6-36 months) (Lowe-McConnell, 1982; Philippart and Ruwet, 1982) and brooding period can be as little as ~ 2 weeks (Coe, 1966). Generation time is sometimes defined as the mean age at which individuals reproduce in a population (Bakker *et al.*, 2022). The average sexual maturation time is therefore likely to occur much earlier than the generation time, especially in a lekking species like *Alcolapia,* where older and larger individuals may produce the majority of the offspring. The lack of reliable generation time estimates among cichlids and other taxa remains problematic for scaling these divergence times (Svardal *et al.*, 2021; Bakker *et al.*, 2022). Nonetheless, even given a conservatively short generation time (e.g. 3 months), estimates for the divergence time of the *Alcolapia* most recent common

ancestor still predates the Natron-Magadi split and the connected southern river (Roberts *et al.*, 1993; Dommain *et al.*, 2022).

The substitution rate is based on a point estimate for Malawi cichlids and has associated uncertainty (95% CI 1.6×10^{-9} to 4.6×10^{-9}) (Malinsky et al., 2018a). Kavembe *et al.* (2016) used the default human substitution rate of 2.5×10^{-8} for *A. grahami* demographic models, almost an order of magnitude higher than the substitution rate we used. Some have suggested that *Alcolapia* may have elevated mutation rates compared to other cichlid radiations owing to the hostile conditions and high ultraviolet light levels (Seegers et al., 2001; Wilson et al., 2004; Pörtner et al., 2010). However, if the extreme conditions have raised mutation rates, this may have only come about in the last 7 ka when soda conditions emerged and not over the entire course of their speciation (Roberts et al., 1993).

Our estimates are also limited by the incomplete sampling of all the *Alcolapia* populations and the omission of *A. ndalalani* samples. Insufficient sampling, low sample sizes and genotype accuracy all affect SFS estimations (Excoffier *et al.*, 2013; Terhorst and Song, 2015). We also restricted the types of demographic changes to bifurcating splits, continuous bidirectional gene flow and instantaneous population size changes. In reality, the emergence of a new deme is not always a straight split (e.g. hybrid speciation), and admixture and growth rates vary over time. We also did not consider the existence of “ghost populations”, which are demes that went extinct in the past and can have a significant effect on SFS estimates (Ottenburghs, 2020). Consequently, we caution against the interpretation of the results given the constraints from the parameters provided. Nonetheless, there are practical limits to increasing the complexity of the parameter space, which becomes greater the smaller the sample size (Excoffier *et al.*, 2013).

CHAPTER 5: GENERAL DISCUSSION

5.1 Research findings

In the work presented in this thesis, both the patterns of diversification and mechanisms of speciation in the *Alcolapia* species flock were investigated. By employing an integrative approach, we focused on three key aspects of speciation: assortative mate choice, ecological niche diversification, and past demographic processes. A range of molecular approaches was utilised to investigate these mechanisms, from the use of microsatellite markers to perform paternity analyses to metabarcoding for diet and microbiome characterisation and the use of genomic datasets to perform population genetic analyses and demographic inference.

In Chapter 2, the barriers underlying *Alcolapia* reproductive isolation were explored by characterising the degree of assortative mating between species. Variable levels of assortative mating between sympatric species suggest reproductive isolation is incomplete, and this aligns with the evidence of ongoing gene flow demonstrated in Chapter 4. Moderate to strong assortative mating by *A. grahami* suggests assortative mating could emerge or be maintained in allopatry.

In Chapter 3, we focused on the ecological aspects of speciation. A metabarcoding approach was used to identify cryptic diversity in diets and gut microbial communities, which was not previously possible using traditional methods. This revealed considerable algal diversity in the *Alcolapia* diet, although many taxa were poorly resolved, indicating an absence of described and sequenced taxa in reference databases. Echoing previous findings, we found sympatric species to be largely algivorous, with high dietary overlap and only minor species differences. These results were indicative of dietary niche segregation primarily through alternative foraging modes in addition to some exploitation of different resource types. Differences in the abundance of some dietary resources (e.g. midges) between sites and sampling years suggest that the availability of trophic resources varies both spatially and temporally.

Compared to other cichlid radiations, the most abundant microbial genera were more often associated with extreme conditions, suggesting an environmental origin of many gut microbes. We also found evidence that *Alcolapia* lack some of the core microbes

present in cichlids from other radiations, potentially in response to the extreme conditions. *Alcolapia* species exhibited variation in their microbial communities that was more consistent than the phototrophic composition findings, suggesting that the microbiome may better reflect species differences. Using these approaches, we could not distinguish between gut-living and environmentally-derived microbes; therefore, we can only speculate on the relative contribution of host physiology and diet to gut communities.

In Chapter 4, we used genomic data to quantify patterns of gene flow between species and sub-populations at a greater level of detail than in previous studies. We found that relative rates of gene flow varied between sympatric species and within sub-populations. Demographic modelling revealed the paraphyly of the Natron radiation and supported an origin for *Alcolapia* much earlier than the formation of Lakes Natron and Magadi. We also found that gene flow has likely occurred throughout the course of *Alcolapia*'s speciation, including significant admixture between ancestral lineages.

5.2 On the boundaries of speciation: multiple mechanisms acting in concert

The rapid radiations of cichlids, where distinct species can arise without significant spatial and ecological barriers, have been of keen research interest to evolutionary biologists (Kocher, 2004; Dieckmann *et al.*, 2012). The Natron *Alcolapia* radiation takes this to the extreme, with three morphologically divergent species coexisting within small and shallow habitats lacking in niche complexity (Seegers and Tichy, 1999; Ford *et al.*, 2016). Our findings of partial assortative mating (Chapter 2), extensive dietary niche overlap (Chapter 3) and ongoing gene flow (Chapter 4) all demonstrate that reproductive isolation is significant but remains incomplete in this system. Similar to most cichlids, the *Alcolapia* species designations are not consistent with the biological species concept (Kocher, 2004), although they are considered legitimate species following the morphological, genetic clusters and cohesion species concepts (Templeton, 1989; Mallet, 1995; Ford *et al.*, 2015; Fricke *et al.*, 2022). Nonetheless, their partial isolation indicates that they are still in the early stages of speciation.

5.2.1 The role of assortative mating and ecological partitioning in *Alcolapia* speciation

The divergent trophic morphology, dietary segregation, and partial assortative mating suggest that both ecologically and sexually mediated processes are involved in reproductive isolation between sympatric species of *Alcolapia*. This is consistent with theoretical and empirical studies, which found that speciation in sympatry/parapatry is more likely when there is a combination of ecological diversification and assortative mate choice (Bolnick and Fitzpatrick, 2007; Ritchie, 2007; Van Doorn *et al.*, 2009; Safran *et al.*, 2013; Scordato *et al.*, 2014). While assortative mating is likely the primary reproductive barrier, ecomorphological and dietary niche differentiation among sympatric *Alcolapia* species suggests that sexual selection may be insufficient for maintaining reproductive isolation. On the other hand, the allopatric *Alcolapia* species exhibited assortative mate choice (Chapter 2) while possessing similar trophic morphology (Ford *et al.*, 2016), suggesting that assortative mating might emerge in allopatry without major ecological diversification. Unlike between ecologically divergent Natron species, these barriers may break down in a secondary contact scenario given their shared generalist trophic morphology with *A. alcalica*.

A key remaining question is whether sexual and ecological selection directly interact to promote speciation in *Alcolapia*. Theoretical models and empirical studies support sympatric speciation when phenotypes conferring reproductive isolation and ecological adaption are genetically linked, either directly through the same genes or indirectly through linkage disequilibrium (Dieckmann *et al.*, 2012; Foote, 2018). These adaptive phenotypes have been referred to as “magic traits” when the same loci are involved or “pseudomagic traits” in the case of tightly linked genotypes (Maan and Seehausen, 2011; Servedio *et al.*, 2011; Servedio & Bürger, 2020). An example is found in Darwin’s finches (*Geospiza* spp.), where beak morphology affects both trophic partitioning and song involved in mate preferences (Podos, 2001; Huber *et al.*, 2007). In *Alcolapia*, male colouration differences between species could be related to interspecific dietary differences, but this is unlikely as aquarium-reared specimens had similar colouration to wild specimens given identical diets (*personal observation*). Moreover, some trophic morphological characters are used as mating signals; for instance, the enlarged lips of *A. latilabris* may facilitate both algae scraping and mate

choice. This has been documented in Midas cichlids (*Amphilophus* spp.) possessing hypertrophied lips (Machado-Schiaffino *et al.*, 2017).

In other contexts such as in sensory drive, sexual selection alone can operate on mating signals and preferences, yet ecological conditions influence the expression of these sexual signals (Safran *et al.*, 2013). This phenomenon is probably less common than when ecological selection acts independently on a trait also subject to sexual selection (Maan and Seehausen, 2011; Scordato *et al.*, 2014; Mendelson and Safran, 2021) and seem unlikely to be a major feature among current *Alcolapia* populations due to the shallow habitat, which precludes the existence of multiple light environments. Alternatively, this may have played a role in the deeper (~55M) paleolake (Seehausen *et al.*, 2008; Dommain *et al.*, 2022).The study of the gut microbiome and its role in hosts' evolution is still a nascent field, but initial studies suggest that gut microbial communities have the potential to influence both sexual and ecological traits involved in speciation (Shropshire & Bordenstein, 2016; Miller *et al.*, 2021). Our finding that *Alcolapia* species differ in their gut microbial communities raises the prospect of microbial-influenced speciation. At this stage, studies largely cannot identify the link between microbial communities and other ecological and sexual traits (Rennison *et al.*, 2019; Groussin *et al.*, 2020; Miller *et al.*, 2021). A key question will be how much of a role microbial diversification plays in speciation relative to other speciation mechanisms.

5.2.2 The role of past processes in *Alcolapia* speciation

Hybridisation between ancestral lineages has been a major driving force in adaptive radiations (Berner and Salzburger, 2015; Jamie and Meier, 2020). For instance, in *Heliconius* butterflies, hybridisation has contributed to the diversification of wing patterns involved in mimicry and mating signals (Wallbank *et al.*, 2016; Enciso-Romero *et al.*, 2017). Recent research has found that hybridisation and the retention of ancestral polymorphisms have been particularly important in providing genetic diversity for driving cichlid radiations, both within the African great lakes (Meier *et al.*, 2017a; McGee *et al.*, 2020; Ronco *et al.*, 2020; Svardal *et al.*, 2021) and in smaller radiations (Kautt *et al.*, 2016; Poelstra *et al.*, 2018). Likewise, past periods of allopatry have promoted speciation in cichlids (Sturmbauer *et al.*, 2001; Genner and Turner, 2005; Joyce *et al.*, 2011).

Divergence times for the *Alcolapia* common ancestor indicate that initial divergence predated the formation of Natron and Magadi as distinct waterbodies (Chapter 4; Dommain *et al.*, 2022). High levels of gene flow estimated between ancestral lineages suggest a similar process may have occurred in *Alcolapia*, whereby past hybridisation may have influenced recent speciation. Given the estimated timeline, multiple hydrological and geographic changes likely occurred during divergence (Owen *et al.,* 2018; Dommain *et al.*, 2022). Since we only modelled continuous gene flow, we cannot rule out some past periods of allopatry, a prerequisite for sympatric speciation. On the other hand, we omitted *A. ndalalani* in demographic models, and it is still possible that *A. latilabris* and *A. ndalalani* emerged sympatrically in Lake Natron.

5.2.3 Multiple processes acting in concert?

Our findings of assortative mating and dietary niche overlap are indicative of multifarious selection driving speciation in *Alcolapia* (Nosil, 2012). Studies investigating the process of adaptive radiation across different taxa have described a pattern of divergence that occurs in stages. This first starts with macrohabitat segregation, then with trophic niche specialisation, and is followed by divergence in signalling corresponding to species recognition and mate choice (Streelman and Danley, 2003; Gavrilets and Losos, 2009). Support for this form of adaptive radiation is found in Tanganyikan cichlids (Ronco *et al.*, 2020), while diet and macrohabitat divergence likely occurred concurrently in Lake Victoria cichlids (McGee *et al.*, 2020). As these analyses rely on parallel processes between multiple species, an analogous investigation into the stages of *Alcolapia* speciation is not feasible, but given the rapid nature of the radiation, they likely occurred either simultaneously or in quick succession.

Overall, the results of this thesis suggest that *Alcolapia* speciation is driven by multiple processes acting in concert. The absence of multiple species at many sites in northern Natron and throughout Magadi suggests that while mate choice may be the primary barrier, ultimately, their speciation may be contingent on the ecological opportunity provided within the wider habitat. This restricted distribution of sympatric species suggests the strength of reproductive barriers between Natron species is likely to be close to the threshold necessary for species to stably coexist.

An open question is whether historical hybridisation promoted the emergence and persistence of sympatric species, and whether other taxa or extinct ghost lineages transferred genes with *Alcolapia* in the past. Given the pervasiveness of this phenomenon in both small and large cichlid radiations alike (Brawand *et al.*, 2015; Meier *et al.*, 2017a; Malinsky *et al.*, 2018a; Poelstra *et al.*, 2018; Kautt *et al.*, 2020; McGee *et al.*, 2020; Ronco *et al.*, 2020), it seems probable that alongside assortative mating and ecological diversification, ancestral hybridisation has also influenced *Alcolapia* speciation.

5.3 Future directions

Our findings suggest that the *Alcolapia* system involves many speciation processes found in larger cichlid adaptive radiations, but on a much smaller spatial and temporal scale. The findings from this thesis open up new possibilities or analyses, including the genetic characterisation of *Alcolapia* adaptations and species phenotypes.

Testing the degree of reproductive isolation between different populations could provide insights into the order that reproductive barriers emerged across different spatial scales and along different stages of the speciation continuum. For instance, premating isolation could be quantified between populations of northern allopatric and southern sympatric *A. alcalica*, between *A. grahami* from separate lagoons, and between yellow and blue sympatric morphs. Aquarium mating trials could be complemented through measurements of reproductive isolation in the field. The collection and microsatellite genotyping of mouthbrooding females and their offspring is one such approach for inferring assortative mating (Martin, 2013). Using this method, comparisons between different sub-sites could be used to infer how levels of assortative mating vary depending on different environmental factors, such as between spring sites and lagoons or transparent versus turbid waters (Seegers and Tichy, 1999).

One potential area of investigation is to characterise the phenotypes involved in mate choice, which can be achieved by isolating respective mating cues with the use of laboratory manipulations. In cichlids, altering the light spectrum has been used

effectively to eliminate male colouration differences (e.g. Seehausen and Van Alphen, 1998; Selz *et al.*, 2014). A similar approach may not be feasible in *Alcolapia* as male colouration occurs in more indiscrete patches that vary between species. In addition, cues involved in mate choice include olfactory compounds, male behaviours and the shape of bowers or courtship structures, which could all be investigated with the use of field or laboratory manipulations (Plenderleith *et al.*, 2005; Genner *et al.*, 2008).

While prezygotic mate choice is likely the primary reproductive barrier in *Alcolapia* (Rometsch *et al.*, 2020), other postzygotic barriers may contribute to reproductive isolation. Generating hybrid crosses and measuring additional factors such as embryonic development, growth and fertility rates could be used to estimate intrinsic postzygotic barriers. In addition, providing parental species with a choice of either hybrid or conspecific males could be used to explore behavioural postzygotic barriers.

The underlying genetic basis of *Alcolapia* species' phenotypes could be characterised using QTL (Quantitative Trait Loci) mapping, where F2 hybrids are generated by reciprocal crosses, phenotypes are quantified, and samples are sequenced (Turner *et al.*, 2013). QTL mapping allows the statistical inference of gene regions controlling phenotypes. In *Alcolapia,* this could be particularly powerful if male colouration, trophic morphology and female choice phenotypes are quantified. In particular, this technique could be used to discover if speciation genes have a simple or polygenic genetic basis or occur in linked regions previously identified as highly differentiated (Turner *et al.*, 2005; Ford *et al.*, 2015; Kautt *et al.*, 2020). An alternative or complementary approach could be to perform GWAS (Genome-wide Association Studies) of naturally admixed populations exhibiting polymorphism (Bush and Moore, 2012; Santos *et al.*, 2023). These methods could also be used to investigate the genes controlling colouration in sympatric blue and yellow male *A. alcalica* morphs, which do not cluster separately based on RADseq and whole genome data (Chapter 4; Ford *et al.,* 2015). The use of gene-editing technologies such as CRISPR-Cas9 could be used to manipulate the genotypes controlling species traits, providing further validation of these findings (Clark *et al.*, 2022; Santos *et al.*, 2023).

Phenotypic plasticity can promote diversification by facilitating phenotypic divergence into different niches (Pfennig and Pfennig, 2009). Recent studies sequencing the

methylome of closely related species suggest that epigenetic changes may facilitate the occupation of ecological niches more rapidly than genomic evolution (Vernaz *et al.*, 2022). Aquarium populations of *A. latilabris* exhibit significantly reduced lip sizes in comparison to wild populations (*personal observation*). This suggests potential plasticity in this trait that may be mediated by ecological conditions in the wild, such as from algae-scraping on abrasive surfaces. A similar phenomenon has been experimentally tested using aquarium populations of thin- and thick-lipped Nicaraguan Midas cichlids (Machado-Schiaffino *et al.*, 2014). By manipulating the feeding surface (using sandpaper and volcanic rocks), lip area was significantly larger in treatment groups of the thick-lipped species only (Machado-Schiaffino *et al.*, 2014). Increasingly, studies have also used transcriptomics to identify candidate genes underlying species traits (Baldo *et al.*, 2011; Santos *et al.*, 2023). Transcriptomics has been employed on *A. grahami* to identify genes associated with adaptation to extreme conditions (Kavembe *et al.*, 2015). A similar approach could be applied to Natron species, but to identify morphological and plastic traits such as lip size.

Using metabarcoding of gut contents, many bacteria and algae only resolved to broad taxonomic levels, highlighting the large gaps in reference databases. In the short term, it is impractical for all such taxa to be sequenced and characterised, and therefore a priority could be to focus sequencing efforts on the most abundant but poorly resolved cyanobacteria taxa. The short reads generated from amplicon sequencing are also limited in resolving fine-scale diversity. Future studies could use a metagenomic approach, where all the DNA within a sample is sequenced. This approach can also extend beyond taxonomical classification and allow for the functional characterisation of gene families (Mirete *et al.*, 2016). A complement of approaches can be particularly effective for characterising both community and functional differences in the microbiome. For instance, Greene *et al.* (2020) combined metabarcoding, metagenomics and metabolomics to investigate microbiome diversification in Lemurs. In addition to differences in diversity and composition, they discovered that mature leaf-eating specialists had more taxa and metabolites associated with cellulolysis than more generalist species (Greene *et al.,* 2020).

In Chapter 3, a further issue was the inability to distinguish between environmental and resident gut microbes. Extraction of environmental DNA (eDNA) from water

samples and substratum could be used to classify microbial taxa of environmental origins. While eDNA water samples were collected on the previous 2017 collection trip, in our metabarcoding study, we were unable to yield reliable data due to issues with contamination. Additionally, intrinsic species differences in the gut microbiome could be explored using aquarium populations fed on an identical diet, eliminating any differences based on diet or habitat partitioning (e.g. Baldo *et al.*, 2015).

While the diet and microbiome of Natron *Alcolapia* have now been characterised, habitat partitioning has only been recorded anecdotally (Seegers and Tichy, 1999; G.F. Turner *personal communication*, 16th July 2022). Interspecific habitat division could be investigated by characterising microhabitat types (substrate, vegetation, flow, water chemistry etc.) and recording the prevalence of different species at these microhabitats. Alternatively, habitat selection experiments could be carried out in aquarium populations by providing a choice of different habitat types. In addition, competitive exclusion by species could be investigated with the use of aquarium experiments (e.g. Winkelmann *et al.*, 2014).

In Chapter 4, we did not include *A. ndalalani* in demographic scenarios due to computational and time restraints, but a clear next step for further research would be to infer the divergence time of *A. ndalalani* relative to the other species splits. Demographic inference using the whole genome dataset could also complement the findings from the RADseq dataset. A number of methods are not based on the allele frequency spectrum but use many neutrally evolving loci (e.g. BPP and G-PhoCS) and therefore do not require large sample numbers (Gronau *et al.*, 2011; Flouri *et al.*, 2020).

Based on the extensive gene flow and incomplete lineage sorting in *Alcolapia*, their phylogenetic relationships may be better modelled using a network approach as opposed to tree-based methods (Morrison, 2014). We attempted phylogenetic network inference using both *PhyloNet* and *PhyloNetworks,* but these yielded inconsistent results, likely as a result of an incorrect input species tree and due to the highly reticulated nature of *Alcolapia* relationships (Than *et al.*, 2008; Zhu *et al.*, 2018; Olave and Meyer, 2020).

The whole genome re-sequenced dataset used in Chapter 4 was not suitable for many types of analysis due to the low sequencing depth and small sample sizes. A higher coverage whole genome dataset could allow for more robust demographic inference and analyses, which require higher genotype confidence, such as genome scans (Nosil and Feder, 2012). Likewise, a more comprehensive whole genome dataset consisting of multiple samples per population would enhance the precision of these demographic estimations. In addition, new approaches such as “haplotagging”, a potentially low-cost method that preserves haplotype information, promise to improve the accuracy of ancestral reconstruction and extend to other analyses, such as identifying chromosomal rearrangements (Meier *et al.*, 2021).

Divergence time inference could also be improved with a more accurate substitution rate and generation time estimates. The per-generation mutation rate could be inferred by high coverage (>40x) sequencing of parent and offspring trios (Malinsky *et al.*, 2018a). To maintain the effect of environmental conditions, this could be carried out by cordoning off wild gravid females with individual males and collecting the resulting embryos. Estimates of generation time could be achieved using a mark and recapture approach, where females' growth rate and egg production are measured over time and used as a proxy to infer generation time (Beverton, 1992; Jonasson *et al.*, 2022).

In addition to genomic analyses, a more comprehensive fossil collection could help to elucidate the historical speciation of *Alcolapia*. In particular, expanding the fossil collections to encompass a broader range of locations surrounding Lake Natron could help reveal the timing of origin for the more derived downturned mouth morphology found in *A. latilabris* and *A. ndalalani.*

5.4 Conservation and threats

Alcolapia are locally abundant at many sites, yet their highly restricted distribution and fragile habitat mean that all *Alcolapia* are listed as vulnerable or endangered (Bayona, 2006; Bayona & Akinyi, 2006). The presence of hybrid sites and single species sites suggests that sympatric species may be vulnerable to the breakdown of reproductive barriers and the reversal of speciation given sufficient environmental changes (e.g. eutrophication). Also, sites in the north-eastern parts of Natron are yet to be sampled

due to their remoteness, which potentially contains undiscovered morphological and genetic diversity (Seegers and Tichy, 1999; Tichy and Seegers, 1999; Ford *et al.*, 2015). Due to their limited genetic diversity, *Alcolapia* populations are potentially at risk of extinction if inbreeding rates rise and they fail to adapt to environmental perturbations (Ford *et al.*, 2015; Svardal *et al.*, 2021).

Alcolapia species face many anthropogenic threats, including climate change. The mean air temperature of the Magadi-Natron basin is projected to increase 2.5 °C by 2050, which may reduce water levels at some sites or cause water temperatures to rise above the thermal limits for survival (Kalacska *et al.*, 2017). In addition, the shallow depths of the spring sites and lagoons mean they are particularly susceptible to changes in groundwater levels (Kalacska *et al.*, 2017).

Maina *et al.* (2019) recorded skewed sex ratios and poor condition factors of multiple *A. grahami* populations, indicating their potential demise. Extraction of water for irrigation near the lakes, human traffic to the lagoons for recreational and health benefits, deforestation and the introduction of parasites are growing threats to Magadi *Alcolapia* populations. Development for a soda ash mine was considered in Natron, but the costs to ecotourism and wildlife were found to outweigh the economic benefits (Kadigi *et al.*, 2014; Shoo, 2020). As the area opens up to more ecotourism, the long-term monitoring of populations and their habitats is needed to safeguard their future.

APPENDIX

Table A2.1. Summaries of all the commands and model outputs the GLMMs and GLMs used in the mate choice analysis.

		Command		Coefficient	Estimate	Standard error	*z* value	*p* value
Brood size	Allopatric	glmer(brood_size ~ female_species + (1|female_id), family="poisson")		(Intercept)	3.2262	0.1967	16.399	<2e-16
				female_speciesGRAHAMI	-0.4997	0.2697	-1.853	0.0638
	Sympatric	glmer(brood_size ~ female_species + (1|female_id), family="poisson")		(Intercept)	2.98004	0.09752	30.56	< 2e-16
				female_speciesLATILABRIS	-0.73321	0.15085	-4.861	1.17E-06
				female_speciesNDALALANI	-0.26306	0.13881	-1.895	0.0581
Assortative courtship time (time spent courting)	Allopatric	glmmTMB(n ~ male_species * female_species + (1|male_id) + (1|date) + (1|time_of_day), family = ziGamma(link = "log"), ziformula=~ species * female_species)	Conditional model	(Intercept)	2.6991	0.1734	15.568	< 2e-16
				male_speciesgra	-1.065	0.2212	-4.815	1.47E-06
				female_speciesGRAHAMI	-0.8932	0.1806	-4.945	7.63E-07
				male_speciesgra:female_speciesGRAHAMI	1.9009	0.2539	7.486	7.10E-14
			Zero-inflation model	(Intercept)	-1.5755	0.2589	-6.085	1.17E-09
				male_speciesgra	1.1701	0.3267	3.581	0.000342
				female_speciesGRAHAMI	1.3266	0.3252	4.08	4.51E-05
				male_speciesgra:female_speciesGRAHAMI	-1.6575	0.4346	-3.813	0.000137
	Sympatric	glmmTMB(n ~ male_species * female_species + (1|male_id) + (1|date) + (1|time_of_day), family = ziGamma(link = "log"),	Conditional model	(Intercept)	2.1499	0.3995	5.382	7.38E-08
				male_speciesLAT	-0.2775	0.4839	-0.573	0.56633

		ziformula=~ species * female_species)		male_speciesNDA	-1.2945	0.4851	-2.669	0.00762
				female_speciesLATILABRIS	-0.3239	0.511	-0.634	0.52619
				female_speciesNDALALANI	-0.0212	0.4117	-0.052	0.95892
				male_speciesLAT:female_speciesLATILABRIS	1.2675	0.5817	2.179	0.02935
				male_speciesNDA:female_speciesLATILABRIS	0.8276	0.5963	1.388	0.16516
				male_speciesLAT:female_speciesNDALALANI	-0.4155	0.514	-0.808	0.41883
				male_speciesNDA:female_speciesNDALALANI	2.0447	0.5037	4.059	4.93E-05
			Zero-inflation model	(Intercept)	1.8718	0.3101	6.036	1.58E-09
				male_speciesLAT	-1.2771	0.3803	-3.358	0.000785
				male_speciesNDA	-0.8177	0.3934	-2.078	0.037681
				female_speciesLATILABRIS	0.2076	0.4568	0.455	0.649416
				female_speciesNDALALANI	-0.4146	0.4107	-1.009	0.312776
				male_speciesLAT:female_speciesLATILABRIS	-0.8468	0.5492	-1.542	0.123086
				male_speciesNDA:female_speciesLATILABRIS	-0.2654	0.5695	-0.466	0.641144
				male_speciesLAT:female_speciesNDALALANI	1.0094	0.5285	1.91	0.056123
				male_speciesNDA:female_speciesNDALALANI	-0.662	0.5218	-1.269	0.204504
Assortative courtship frequency (number of courtship attempts)	Allopatric	glmmTMB(courtships ~ female_species * male_species + (1|male_id) + (1|date), family=nbinom2(link = "log"))		(Intercept)	1.2831	0.1602	8.008	1.17E-15
				female_speciesGRAHAMI	-0.7836	0.1621	-4.833	1.35E-06
				speciesgra	-0.7232	0.2277	-3.176	0.00149
				female_speciesGRAHAMI:speciesgra	1.0047	0.2309	4.351	1.36E-05
	Sympatric	glmmTMB(courtships ~ female_species * male_species + (1|male_id) + (1|date) + (1|time_of_day), family=poisson		(Intercept)	-3.3783	0.5362	-6.3	2.97E-10
				female_speciesLATILABRIS	0.4055	0.6222	0.652	0.5146
				female_speciesNDALALANI	0.9163	0.5703	1.607	0.1081
				male_speciesLAT	1.4945	0.6012	2.486	0.0129
				male_speciesNDA	0.9915	0.6328	1.567	0.1172
				male_speciesLAT:female_speciesLATILABRIS	0.4883	0.6781	0.72	0.4714

			male_speciesNDA:female_speciesLATILABRIS	-1.6094	0.6929	-2.323	0.0202
			male_speciesLAT:female_speciesNDALALANI	-0.3102	0.7508	-0.413	0.6795
			male_speciesNDA:female_speciesNDALALANI	0.2776	0.6676	0.416	0.6775
Assortative spawning (brood sired with conspecifics)	Allopatric	glmmTMB(cbind(conspecific,,heterospecific)	(Intercept)	1.4956	0.5362	2.789	0.00529
		~ female_species	mum_speciesNDALALANI	0.8799	0.8997	0.978	0.32809
		family = betabinomial(link = "logit")					
	Sympatric	glmmTMB(cbind(conspecific,heterospecific)	(Intercept)	0.2792	0.6692	0.417	0.677
		~ female_species	female_speciesLATILABRIS	2.0263	1.2539	1.616	0.106
		family = betabinomial(link = "logit")	female_speciesNDALALANI	0.1396	0.8874	0.157	0.875

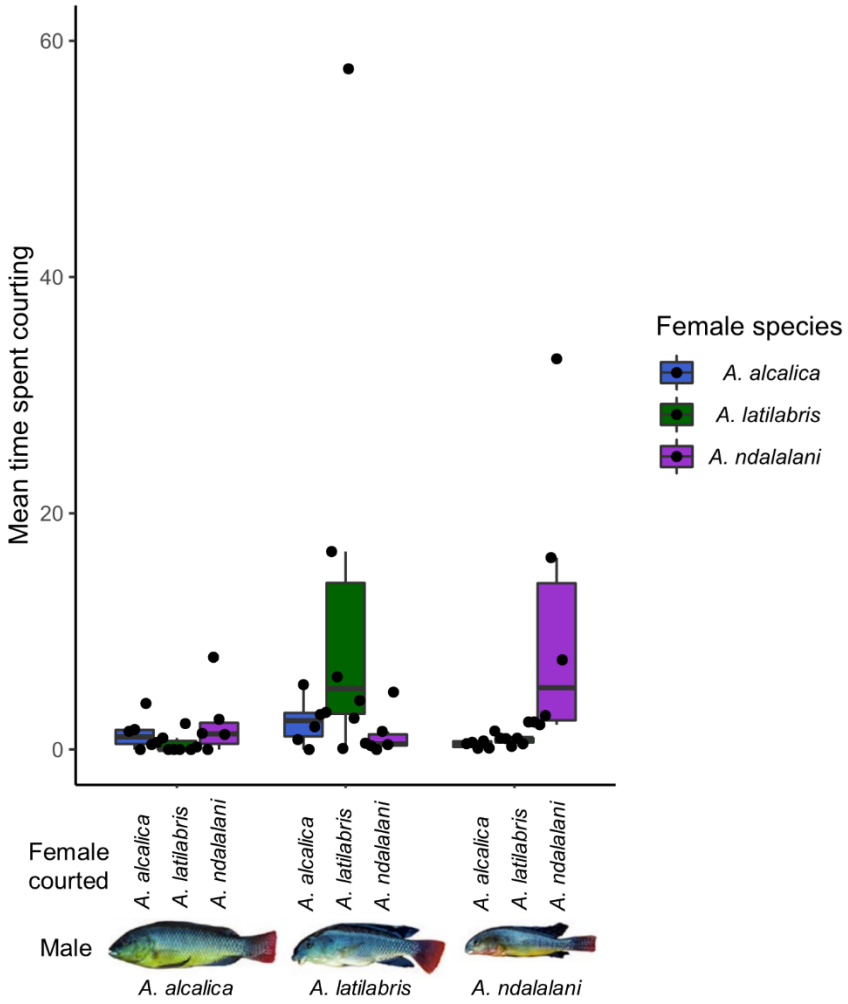


Figure A2.1. Time spent by males performing courtship with conspecific and heterospecific females in sympatric *Alcolapia.* Mean times spent courting for each male are taken from 5-minute focal observations.

Table A2.2. Pairwise post-hoc contrasts of differences in brood size between *Alcolapia* species in allopatric and sympatric experiments. Post-hoc tests are derived using *emmeans* from poisson GLMMs.

	Contrast	**ratio**	**SE**	***z* ratio**	***p* value**
Allopatric	AA♀ vs. AG♀	1.65	0.444	1.853	0.0638
Sympatric	AA♀ vs. AL♀	0.733	0.151	4.861	**<.0001**
	AA♀ vs. AN♀	0.263	0.139	1.895	0.1401
	AL♀ vs. AN♀	-0.47	0.152	-3.094	**0.0056**

Table A2.3. Post-hoc contrasts show differences in time spent in courtship by *Alcolapia* in the allopatric experiment. All estimates and contrasts are derived from post-hoc tests (*emmeans*) from the output of the GLMM. Comparisons are made with for each focal species and their sex, with courtship towards either conspecifics, heterospecifics and both.

		Model estimates						**Pairwise contrast**			
Focal sex	**Focal species**	**Comparison**	**response**	**SE**	***df***	**lower. CL**	**upper.CL**	**ratio**	**SE**	***t* ratio**	***p* value**
Female	AA♀	all ♂	8.73	1.23	408	6.62	11.5	0.944	0.12	-0.452	0.6516
	AG♀		9.24	1.34	408	6.96	12.3				
	AA♀	AA♂	14.87	2.577	408	10.57	20.9	2.901	0.642	4.815	**<.0001**
		AG♂	5.12	0.942	408	3.57	7.36				
	AG♀	AA♂	6.09	1.161	408	4.18	8.85	0.433	0.102	-3.539	**0.0004**
		AG♂	14.04	2.564	408	9.8	20.1				
Male	AA♂	all ♀	9.51	1.51	408	6.97	13	1.12	0.213	0.602	0.5476
	AG♂		8.48	1.36	408	6.19	11.6				
	AA♂	AA♀	14.87	2.577	408	10.57	20.9	2.443	0.441	4.945	**<.0001**
		AG♀	6.09	1.161	408	4.18	8.85				
	AG♂	AA♀	5.12	0.942	408	3.57	7.36	0.365	0.065	-5.66	**<.0001**
		AG♀	14.04	2.564	408	9.8	20.1				

Table A2.4. Post-hoc contrasts show differences in courtship frequency (number of courtship attempts) by *Alcolapia* in the allopatric experiment. All estimates and contrasts are derived from post-hoc tests (*emmeans*) from the output of the GLMM. Comparisons are made with for each focal species and their sex, with courtship towards either conspecifics, heterospecifics and both.

		Model estimates						Pairwise contrast			
		Comparison	response	SE	*df*	asymp. LCL	asymp. UCL	ratio	SE	*t* ratio	*p* value
Female	AA♀	all ♂	2.51	0.3	413	1.99	3.18	1.32	0.152	2.452	**0.0146**
	AG♀	all ♂	1.9	0.231	413	1.49	2.41				
	AA♀	AA♂	3.61	0.578	413	2.63	4.94	2.061	0.469	3.176	**0.0016**
		AG♂	1.75	0.297	413	1.25	2.44				
	AG♀	AA♂	1.65	0.281	413	1.18	2.3	0.755	0.175	-1.215	0.225
		AG♂	2.18	0.362	413	1.58	3.02				
Male	AA♂	all ♀	2.44	0.352	413	1.84	3.24	1.25	0.248	1.112	0.2666
	AG♂	all ♀	1.96	0.286	413	1.47	2.61				
	AA♂	AA♀	3.61	0.578	413	2.63	4.94	2.189	0.355	4.833	**<.0001**
		AG♀	1.65	0.281	413	1.18	2.3				
	AG♂	AA♀	1.75	0.297	413	1.25	2.44	0.802	0.131	-1.353	0.1767
		AG♀	2.18	0.362	413	1.58	3.02				

Table A2.5. Model estimates of time spent courting by sympatric *Alcolapia* with conspecific and heterospecific species. Estimates derive from post-hoc estimations (*emmeans*) of the GLMM. Estimates are provided for each focal species and sex combination with courtship towards conspecifics, heterospecifics and to both species.

	Grouping	Levels	response	SE	df	lower.CL	upper.CL
Overall female courtship time	AA♀	all ♂	5.08	1.12	785	3.3	7.83
	AL♀		7.39	1.69	785	4.72	11.58
	AN♀		8.57	1.83	785	5.63	13.02
Overall male courtship time	AA♂	all ♀	7.65	2.38	785	4.15	14.1
	AL♂		7.7	2.1	785	4.5	13.16
	AN♂		5.46	1.42	785	3.28	9.08
Interspecific courtship time	AA♂	AA♀	2.15	0.399	785	1.366	2.93
		AL♀	1.872	0.313	785	1.259	2.49
		AN♀	0.855	0.32	785	0.227	1.48
	AL♂	AA♀	1.826	0.458	785	0.926	2.73
		AL♀	2.816	0.299	785	2.229	3.4
		AN♀	1.359	0.315	785	0.741	1.98
	AN♂	AA♀	2.129	0.365	785	1.412	2.85
		AL♀	1.436	0.348	785	0.753	2.12
		AN♀	2.879	0.285	785	2.32	3.44

Table A2.6. Post-hoc contrasts show differences in time spent in courtship by *Alcolapia* in the sympatric experiment. Contrasts are derived from post-hoc tests (*emmeans*) from the output of the GLMM. Estimates are provided for each focal species and sex combination with courtship towards conspecifics, heterospecifics and overall (to both species).

	Grouping	Levels	ratio	SE	df	t.ratio	p.value
female courtship propensity (time courting)	AA♀ vs. AL♀	all ♂	0.688	0.148	785	-1.744	0.1894
	AA♀ vs. AN♀		0.593	0.114	785	-2.714	**0.0186**
	AL♀ vs. AN♀		0.863	0.177	785	-0.72	0.7515
male courtship propensity (time courting)	AA♂ vs. AL♂	all ♀	0.994	0.384	785	-0.017	0.9998
	AA♂ vs. AN♂		1.401	0.518	785	0.913	0.6325
	AL♂ vs. AN♂		1.41	0.481	785	1.007	0.5726
male interspecific courtship propensity (time courting)	AA♂	AA♀ vs. AL♀	0.3239	0.511	785	0.634	0.8015
		AA♀ vs. AN♀	0.0212	0.412	785	0.052	0.9985
		AL♀ vs. AN♀	-0.3027	0.455	785	-0.666	0.7834
	AL♂	AA♀ vs. AL♀	-0.9436	0.268	785	-3.522	**0.0013**
		AA♀ vs. AN♀	0.4368	0.303	785	1.442	0.3199
		AL♀ vs. AN♀	1.3804	0.3	785	4.607	**<.0001**
	AN♂	AA♀ vs. AL♀	-0.5037	0.3	785	-1.677	0.2147
		AA♀ vs. AN♀	-2.0235	0.279	785	-7.255	**<.0001**
		AL♀ vs. AN♀	-1.5197	0.276	785	-5.509	**<.0001**
female interspecific courtship propensity (time courting)	AA♀	AA♂ vs. AL♂	0.278	0.484	785	0.573	0.8343
		AA♂ vs. AN♂	1.295	0.485	785	2.669	**0.0212**
		AL♂ vs. AN♂	1.017	0.416	785	2.444	**0.0391**
	AL♀	AA♂ vs. AL♂	-0.99	0.53	785	-1.868	0.1488
		AA♂ vs. AN♂	0.467	0.53	785	0.881	0.6522
		AL♂ vs. AN♂	1.457	0.406	785	3.589	**0.001**
	AN♀	AA♂ vs. AL♂	0.693	0.481	785	1.441	0.3203
		AA♂ vs. AN♂	-0.75	0.436	785	-1.722	0.1975
		AL♂ vs. AN♂	-1.443	0.419	785	-3.444	**0.0017**

Table A2.7. Model estimates of courtship frequency (number of courtship attempts) by *Alcolapia* in the sympatric experiment. Estimates derive from post-hoc estimations (*emmeans*) of the GLMM. Estimates are provided for each focal species and sex combination with courtship towards conspecifics, heterospecifics and overall (to both species).

	Grouping	Levels	response	SE	asymp.LCL	asymp.UCL
female propensity (frequency of courtships)	AA♀	all ♂	0.0781	0.0207	0.0465	0.131
	AL♀		0.1243	0.0298	0.0778	0.199
	AN♀		0.1253	0.0289	0.0797	0.197
male propensity (frequency of courtships)	AA♂	all ♀	0.053	0.0175	0.0278	0.101
	AL♂		0.163	0.0444	0.0951	0.278
	AN♂		0.141	0.0385	0.0828	0.241
Interspecific courtship propensity (frequency of courtships)	AA♂	AA♀	0.0341	0.0183	0.0119	0.0976
		AL♀	0.152	0.0497	0.0801	0.2885
		AN♀	0.0919	0.0349	0.0437	0.1934
	AL♂	AA♀	0.0512	0.0235	0.0208	0.1256
		AL♀	0.3716	0.1027	0.2162	0.6386
		AN♀	0.1011	0.0372	0.0491	0.2081
	AN♂	AA♀	0.0853	0.0328	0.0401	0.1813
		AL♀	0.076	0.0303	0.0348	0.1659
		AN♀	0.3033	0.0855	0.1745	0.5272

Table A2.8. Post-hoc contrasts show differences in courtship frequency (number of courtship attempts) by *Alcolapia* in the sympatric experiment. Contrasts derive from post-hoc estimations (*emmeans*) of the GLMM. Estimates are provided for each focal species and sex combination with courtship towards conspecifics, heterospecifics and overall (to both species).

	Grouping	Levels	ratio	SE	z.ratio	p.value
female courtship propensity (frequency of courtships)	AA♀ vs. AL♀	all ♂	0.326	0.126	-2.89	**0.0108**
	AA♀ vs. AN♀		0.375	0.146	-2.517	**0.0317**
	AL♀ vs. AN♀		1.151	0.398	0.405	0.9134
male courtship propensity (frequency of courtships)	AA♂ vs. AL♂	all ♀	0.628	0.167	-1.748	0.1875
	AA♂ vs. AN♂		0.624	0.161	-1.828	0.1603
	AL♂ vs. AN♂		0.993	0.23	-0.032	0.9994
male interspecific courtship propensity (frequency of courtships)	AA♂	AA♀ vs. AL♀	0.667	0.415	-0.652	0.7914
		AA♀ vs. AN♀	0.4	0.228	-1.607	0.2427
		AL♀ vs. AN♀	0.6	0.299	-1.026	0.5603
	AL♂	AA♀ vs. AL♀	0.409	0.11	-3.314	**0.0027**
		AA♀ vs. AN♀	2	0.787	1.761	0.1828
		AL♀ vs. AN♀	4.889	1.724	4.499	**<.0001**
	AN♂	AA♀ vs. AL♀	0.909	0.382	-0.227	0.972
		AA♀ vs. AN♀	0.303	0.105	-3.438	**0.0017**
		AL♀ vs. AN♀	0.333	0.112	-3.28	**0.003**
female interspecific courtship propensity (frequency of courtships)	AA♀	AA♂ vs. AL♂	0.224	0.1349	-2.486	**0.0345**
		AA♂ vs. AN♂	0.371	0.2348	-1.567	0.2601
		AL♂ vs. AN♂	1.654	0.7784	1.069	0.5335
	AL♀	AA♂ vs. AL♂	0.138	0.0693	-3.938	**0.0002**
		AA♂ vs. AN♂	0.506	0.2838	-1.215	0.4445
		AL♂ vs. AN♂	3.675	1.5705	3.046	**0.0066**
	AN♀	AA♂ vs. AL♂	1.122	0.5869	0.22	0.9738
		AA♂ vs. AN♂	0.281	0.1247	-2.862	**0.0117**
		AL♂ vs. AN♂	0.251	0.1145	-3.03	**0.0069**

Table A2.9. Post-hoc contrasts show differences in the predicted probability of conspecific mating by species in both the sympatric and allopatric experiments. *p* values are testing if the mean expected probability of mating with conspecifics is greater than expected given random mating, at a level of 0.33 for sympatric (3-way choice) and 0.5 for allopatric (2-way choice).

	Species	**Probability of conspecific mating**	**lower. CL**	**upper. CL**	**SE**	***df***	**null**	***t*.ratio**	***p*.value**
Sympatric experiment	AA♀	0.569	0.249	0.841	0.1641	23	0.333	1.453	0.1597
	AL♀	0.909	0.526	0.989	0.0877	23	0.333	2.818	**0.0098**
	AN♀	0.603	0.312	0.836	0.1401	23	0.333	1.899	0.0701
	Overall	0.731	0.51	0.877	0.0913	23	0.333	3.645	**0.0014**
Allopatric experiment	AG♀	0.817	0.591	0.932	0.0802	18	0.5	2.789	**0.0121**
	AA♀	0.915	0.675	0.982	0.0609	18	0.5	3.034	**0.0071**
	Overall	0.874	0.709	0.952	0.0549	18	0.5	3.887	**0.0011**

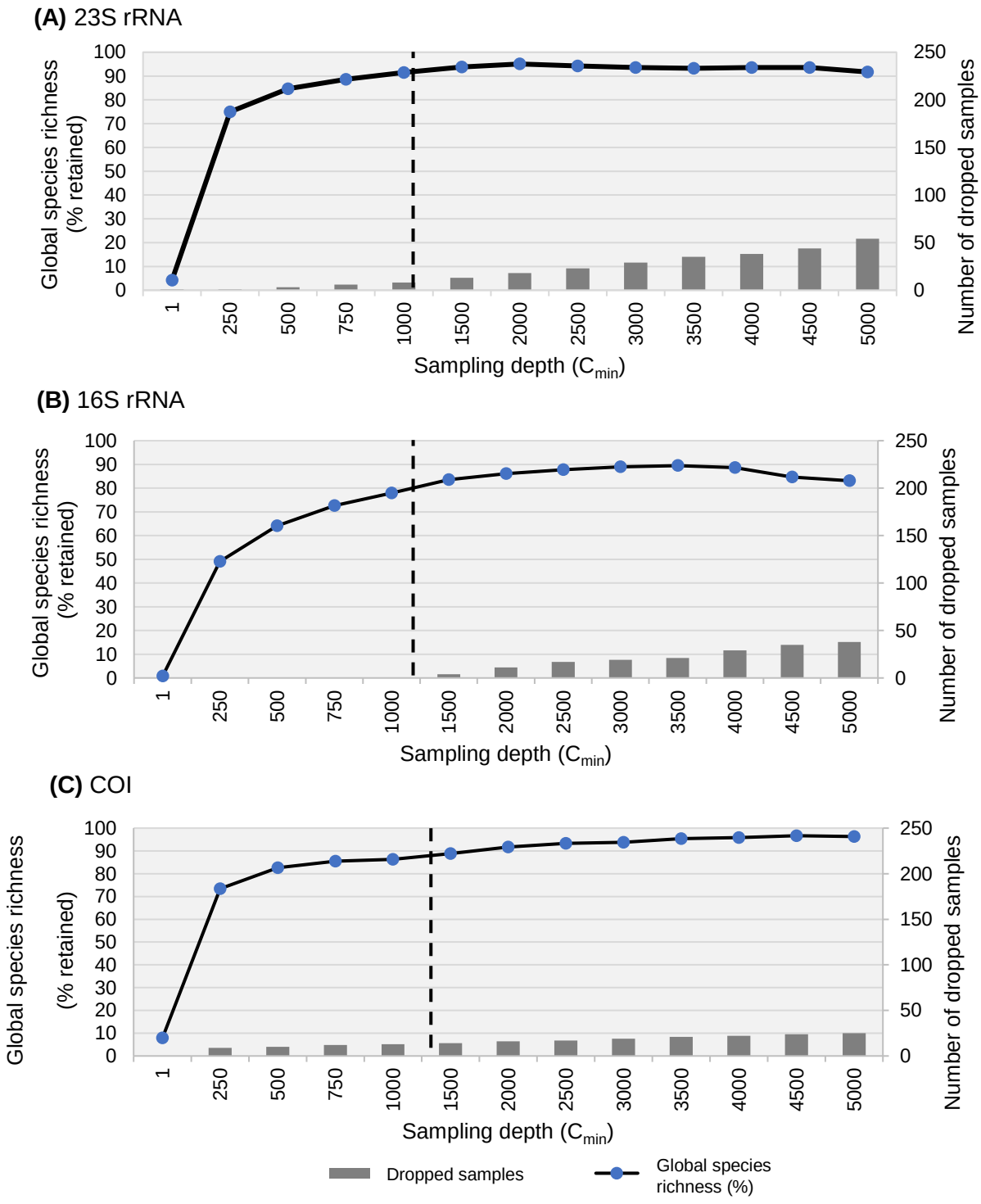


Figure A3.1. Plots showing the effect of changing normalisation value (C_{min}) on global species richness (%) for each amplicon using the program *SRS*. The number of dropped samples for a given C_{min} are show in the grey bars. The dotted line denotes the C_{min} value selected for subsequent analyses. Values are shown for each amplicon after filtering steps and removal of non-target taxa, including cyanobacteria in the 16S bacteria dataset. (**A**) 23S rRNA plastid (targeting: algae and plants) (**B**) 16S rRNA (targeting: bacteria), (**C**) COI (targeting: invertebrates).

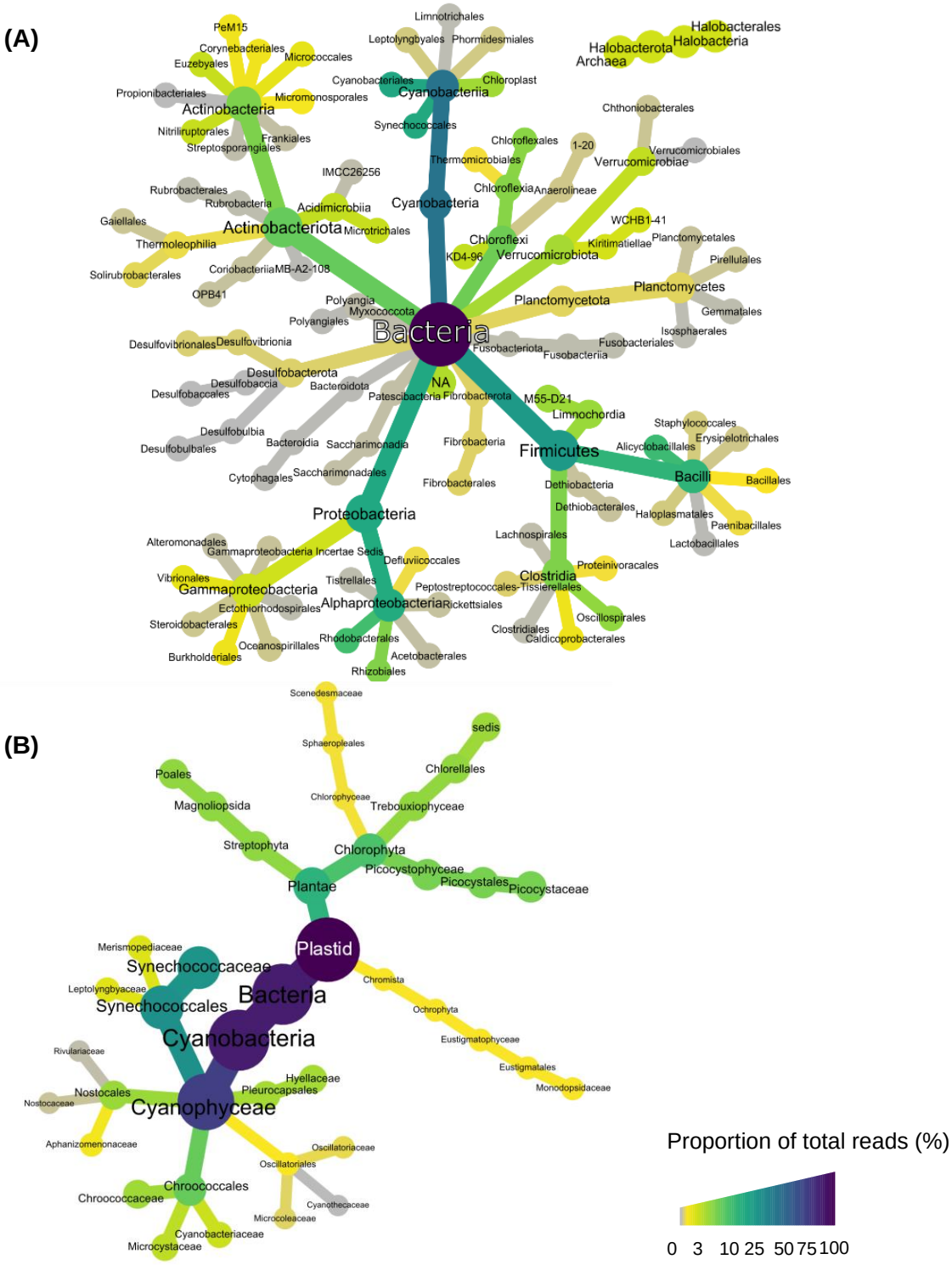


Figure A3.2. The diversity of microbial (16S rRNA amplicon) and phototrophic taxa (23S rRNA) in Natron *Alcolapia* species. Created using metacoder v0.3.5 (**A**) 16S rRNA microbial reads before the removal of cyanobacteria reads (**B**) Phototrophic plastid (23S rRNA) dataset showing the high proportion of cyanobacteria reads. Colour and node size is proportional to relative abundance (total proportion of reads assigned to that taxon). Nodes at the tips of each branch are at the level of Order for microbial sequences (**A**) and at the level of family for the phototrophic sequences **(B).** Orders and families comprising less than 0.01% of the total reads were removed.

Table A3.1. The association with adaptation to extreme conditions of the most abundant microbial genera found in Natron *Alcolapia*. Included are the top 20 genera by mean read abundance (49% of total reads). Dark blue = growth characteristic common to all described species in genus, light blue = growth characteristic common to some described species in genus. Taxa common to both *Alcolapia* and other cichlids are highlighted in red. The mean relative abundance (%) and total number of child taxa (species) listed on LPSN for each genus is provided. *Ca.* = *Candidatus*. Genera traits were assigned by referring to described species using the LPSN database (Carte *et al.,* 2020) and references are provided. * *Methyloceanibacter* species are all found in marine environments, but not confirmed as halophilic in laboratory conditions.

System	Genera	Relative abundance (%)	Total species (LPSN)	Halophilic	Alkaliphilic	Phototrophic	References
Natron *Alcolapia*	***Cetobacterium***	25.2	2				(Foster *et al.*, 1995; Finegold *et al.*, 2003)
	Ca. Chlorothrix	5.2	1				(Klappenbach and Pierson, 2004)
	Rhodobaculum	4.6	1				(Bryantseva *et al.*, 2015)
	Terrimicrobium	3.4	1				(Qiu *et al.*, 2014)
	Egibacter	2.5	1				(Zhang *et al.*, 2016b)
	IMCC26207	1.7	NA				(Kim *et al.*, 2017)
	Salipaludibacillus	1.2	5				(Nielsen *et al.*, 1995; Chen *et al.*, 2009; Sultanpuram and Mothe, 2016; Amoozegar *et al.*, 2018; Wang *et al.*, 2019)
	Sarcina	0.7	5				(Hunter *et al.*, 1981; Goodwin and Zeikus, 1987; Willems and Collins, 1994; Lawson and Rainey, 2016)
	Roseibaca	0.7	1				(Labrenz *et al.*, 2009)
	Bacillus	0.5	622				(Logan and Vos, 2015)
	Ilumatobacter	0.4	5				(Matsumoto *et al.*, 2009, 2013)
	Lactococcus	0.4	26				(Teuber, 1995; Itoi *et al.*, 2008)
	Vibrio	0.3	199				(Farmer *et al.*, 2015)
	Mycobacterium	0.3	263				(Habe *et al.*, 2004; Gupta *et al.*, 2018; Seck *et al.*, 2018)
	ZOR0006	0.3	NA				(Ofek *et al.*, 2021)
	Natronococcus	0.3	5				(Tindall *et al.*, 1984; Kanai *et al.*, 1995; Roh *et al.*, 2007; Corral *et al.*, 2013)
	Alkaliphilus	0.3	12				(Takai *et al.*, 2001; Cao *et al.*, 2003; Wu *et al.*, 2010; Zakharyuk *et al.*, 2017)
	Anaerobacillus	0.3	7				(Zavarzina *et al.*, 2009; Bassil and Lloyd, 2019; Borsodi *et al.*, 2019)
	Egicoccus	0.3	1				(Zhang *et al.*, 2016a)
	Methyloceanibacter	0.3	5	*			(Takeuchi *et al.*, 2014; Vekeman *et al.*, 2016)

Table A3.2. The association with adaptation to extreme conditions of the most abundant microbial genera found in non-*Alcolapia* cichlids (Baldo *et al.,* 2019). Included are the top 20 genera by mean read abundance (48% of total reads). Dark blue = growth characteristic common to all described species in genus, light blue = growth characteristic common to some described species in genus. Taxa common to both *Alcolapia* and other cichlids are highlighted in red. The mean relative abundance (%) and total number of child taxa (species) listed on LPSN for each genus is provided. *Ca.* = *Candidatus*. Genera traits were assigned by referring to described species using the LPSN database (Carte *et al.,* 2020) and references are provided.

System	Genera	Relative abundance (%)	Total species (LPSN)	Halophilic	Alkaliphilic	Phototrophic	References
Other cichlid radiations (Baldo *et al.,* 2019)	***Cetobacterium***	33.6	2				(Foster *et al.*, 1995; Finegold *et al.*, 2003)
	Epulopiscium	2.1	1				(Miyake *et al.*, 2015)
	Plesiomonas	2	1				(Miller and Koburger, 1986; Santos *et al.*, 2015)
	Rhodobacter	1.9	34				(Igeno *et al.*, 1995; Srinivas *et al.*, 2007; Imhoff, 2015)
	Porphyromonas	1.3	23				(Shah, 1992; Guilloux *et al.*, 2021)
	Ca. Rhabdochlamydia	1.2	4				(Kostanjšek *et al.*, 2004; Thomas *et al.*, 2006; Corsaro *et al.*, 2007; Pillonel *et al.*, 2019)
	Clostridium	0.7	326				(Fendrich *et al.*, 1990; Schnürer *et al.*, 1996; Zhilina *et al.*, 2005)
	Luteolibacter	0.6	11				(Jiang *et al.*, 2012; Zhang *et al.*, 2017; Dahal *et al.*, 2021)
	Bacillus	0.6	625				(Matsumoto *et al.*, 2009, 2013)
	u114	0.6	NA				(Kashinskaya *et al.*, 2018)
	Lawsonia	0.5	1				(McOrist *et al.*, 1995)
	Ca. Xiphinematobacter	0.5	5				(Vandekerckhove *et al.*, 2000)
	Pseudomonas	0.4	524				(Fendrich, 1988; Amoozegar *et al.*, 2014; Wong and Lee, 2014; Girard *et al.*, 2021)
	Methylosinus	0.4	3				(Bowman *et al.*, 1993)
	Turicibacter	0.3	1				(Bosshard *et al.*, 2002; Maki and Looft, 2022)
	Dechloromonas	0.3	8				(Achenbach *et al.*, 2001; Horn *et al.*, 2005; Wolterink *et al.*, 2005)
	Shewanella	0.3	96				(Lemaire *et al.*, 2020)
	Anaerospora	0.2	1				(Woo *et al.*, 2005)
	Novispirillum	0.2	1				(Woolley and Woolley, 1987; Yoon *et al.*, 2007; Morishita and Okamoto, 2017)
	Vibrio	0.2	202				(Farmer *et al.*, 2015)

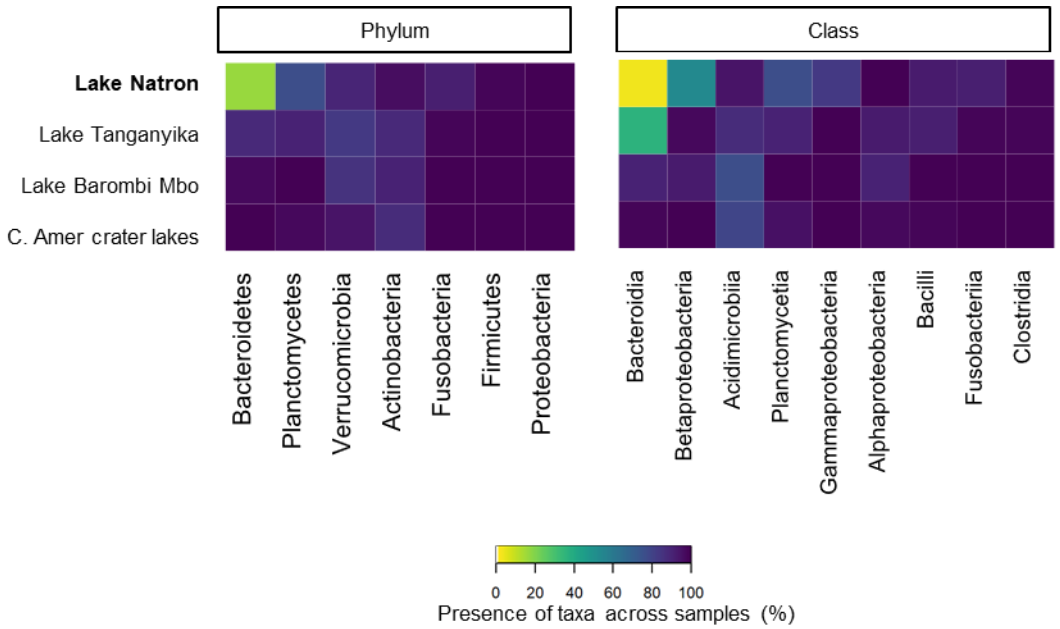


Figure A3.3. The proportion of samples containing some core microbial phyla and classes is lower in Natron *Alcolapia* samples when compared to other cichlid radiations. Core taxa were any taxa that were present in 90% or more of samples from at least one of these four cichlid radiations (Baldo *et al.,* 2019).

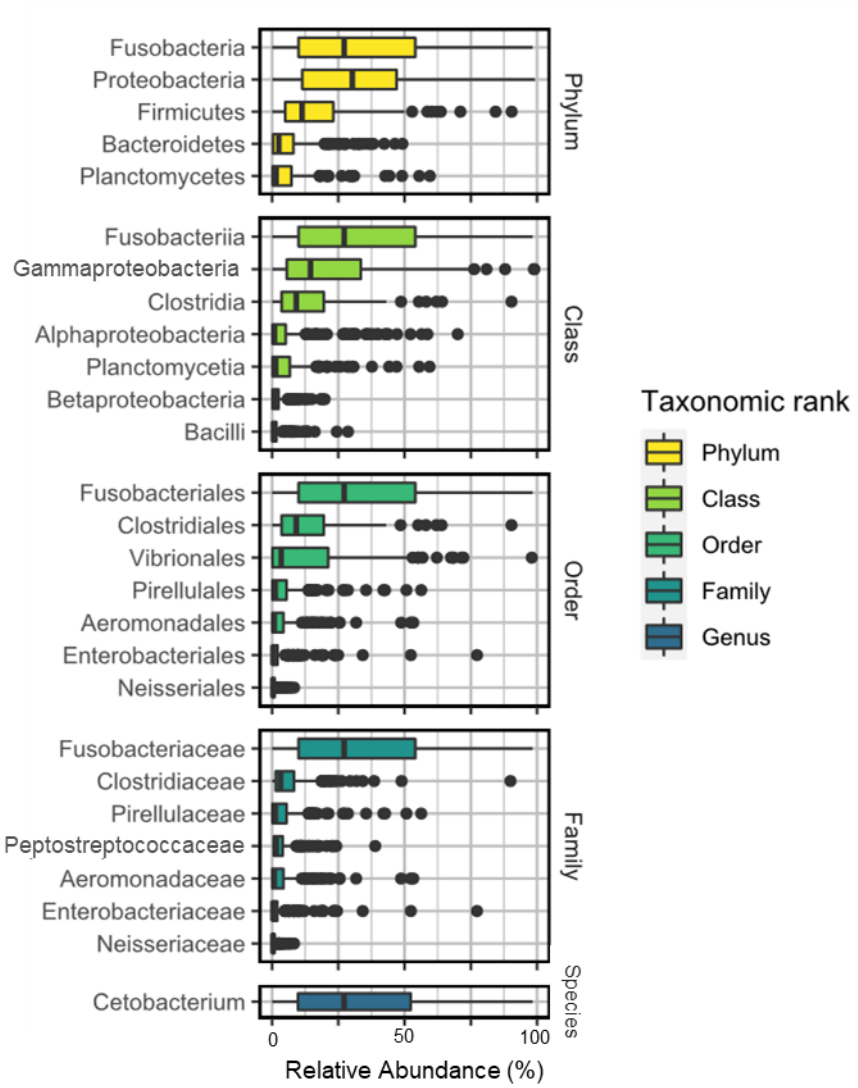


Figure A3.4. The core gut microbiome of other cichlid radiations shown at different taxonomic levels using the OTU dataset from Baldo *et al.,* 2019. Includes taxa present in 90% or more of core taxa samples (>10,000 reads). Includes gut contents of cichlids taken from Lake Tanganyika (Tanzania), Barombi Mbo (Cameroon), Crater lakes (Apoyo, Masaya, Xiloá, Apoyeque, Asosca León) from Nicaragua.

Table A3.3. Sympatric Lake Natron *Alcolapia* species differ in their microbiome (16S rRNA) and dietary phototrophs (23S rRNA) community composition using different distance metrics. R^2 values were generated using PERMANOVA (adonis) with terms of interest. To account for unbalanced sample sizes and heteroscedasity, *p* values and test statistics (T_w^2) were generated for each term using W_d* test (Welch's MANOVA) and strata.

		Microbial Taxa (16S amplicon)						**Phototrophic Taxa (23S Plastid amplicon)**					
		Bray-Curtis			**Generalised Unifrac**			**Bray-Curtis**			**Generalised Unifrac**		
Term	***df***	$\mathbf{R^2}$	$\boldsymbol{T_w^2}$	***p***	$\mathbf{R^2}$	$\boldsymbol{T_w^2}$	***p***	$\mathbf{R^2}$	$\boldsymbol{T_w^2}$	***p***	$\mathbf{R^2}$	$\boldsymbol{T_w^2}$	***p***
Species	2	0.04	2.54	**0.001**	0.05	3.08	**0.001**	0.03	1.95	**0.001**	0.03	1.94	**0.006**
Site	5	0.40	17.69	**0.001**	0.42	18.40	**0.001**	0.56	35.8	**0.001**	0.50	32.0	**0.001**
gut fullness	2	0.02	3.51	**0.007**	0.02	3.74	**0.001**	0.01	3.84	0.09	0.01	3.86	0.068
species:site	7	0.07	7.89	**0.008**	0.08	8.95	**0.001**	0.05	25.4	**0.003**	0.05	17.9	**0.001**
species:gut fullness	4	0.04	2.32	**0.001**	0.03	2.79	**0.001**	0.03	2.34	**0.01**	0.03	17.9	**0.006**
Residuals	100	0.43			0.40			0.33			0.36		
Total	120	1			1			1			1		

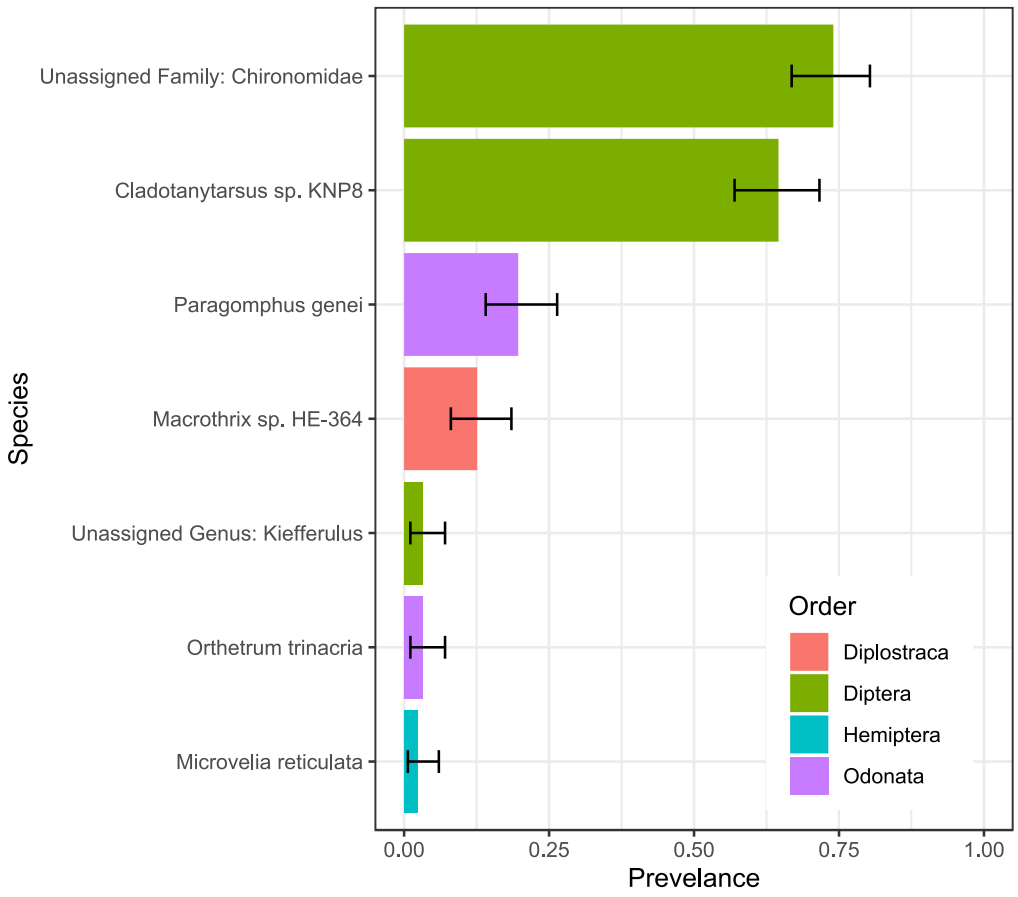


Figure A3.5. The prevalence of different arthropod taxa (COI amplicon) from Lake Natron *Alcolapia* sympatric sites. Where reads were not assigned to an individual species or genus, reads were clustered into the next taxonomic grouping (genus and family). Colours denote the taxonomic Order that the species/cluster belongs to. Error bars are the 95% Clopper-Pearson confidence intervals.

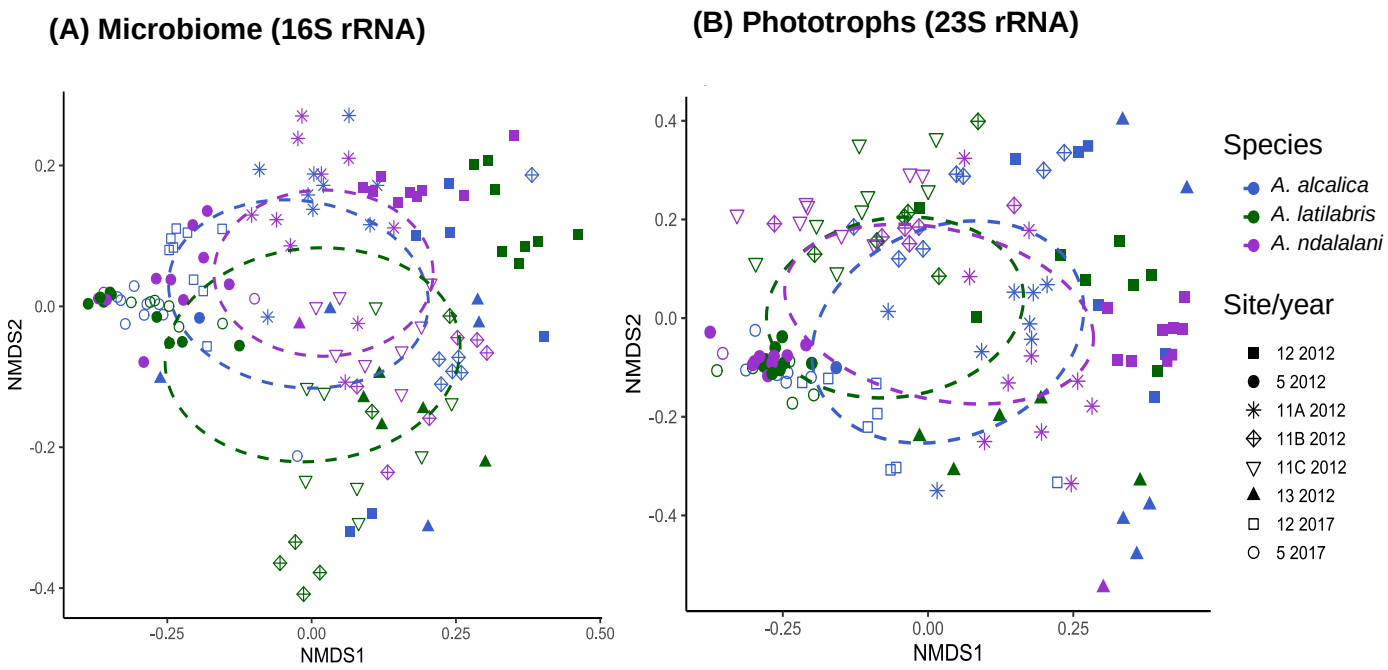


Figure A3.6. NMDS plots showing interspecific differences in community composition of the microbiome and dietary phototrophic taxa in sympatric *Alcolapia* species using generalized UniFrac distance metric. Site differences and species differences in the (**A**) microbiome (16S rRNA) and (**B**) dietary phototrophic (23S rRNA) communities. Ellipses show the 95% confidence interval of the centroid for each species based on the standard errors of average axis scores.

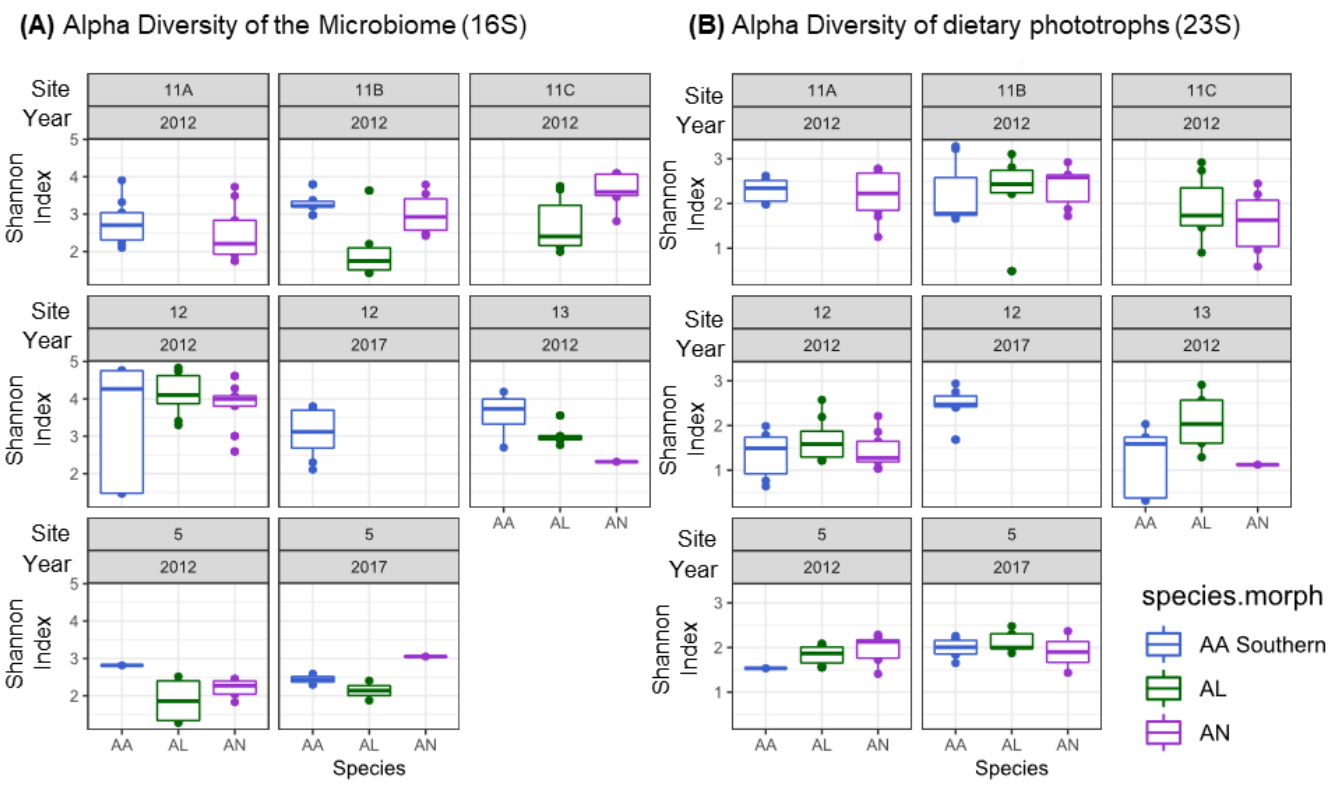


Figure A3.7. Alpha diversity differences between species, sites and years in sympatric *Alcolapia* using the Shannon Index of ASV feature data. Between sympatric species there were no significant differences in the alpha diversity of the (**A**) microbiome (16S rRNA) (GLMM, X^2 = 5.56, p= 0.062) and (**B**) Dietary phototrophic (23S rRNA) taxa (GLMM, X^2 = 2.61, p= 0.27).

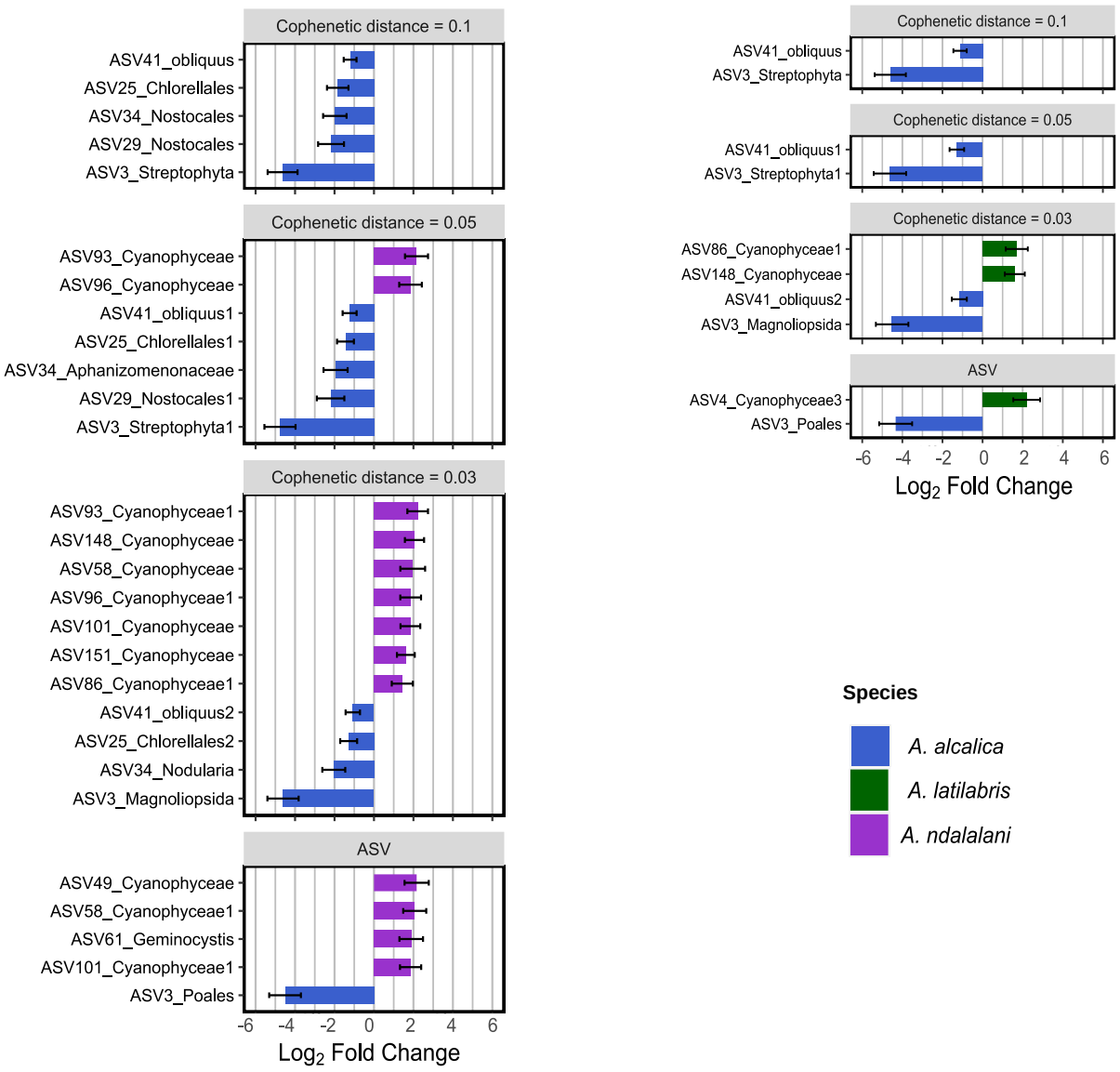


Figure A3.8. Differentially abundant phototrophic taxa (23S rRNA) between *Alcolapia* species sampled from Lake Natron. *A. alcalica* and (**A**) *A. latilabris,* (**B**) *A. ndalalani*, but not between benthic species. Differentially abundant taxa between species were identified using LinDA (Zhou *et al.,* 2022) at 4 different phylogenetic levels. ASVs were clustered by copheneitc distance using the "tip_glom" function in phyloseq, at 3 levels approximating taxonomic level. 0.03 = species or akin to OTU (Operational Taxonomic Unit at 97% similarity), 0.05 = genus, 0.1 = family. ASVs were not agglomerated by higher taxonomic designations (genus, family, order) because many ASVs only assigned to broad taxonomic categories. Only taxa that had a significant (p<0.05) Log_2Fold change are plotted.

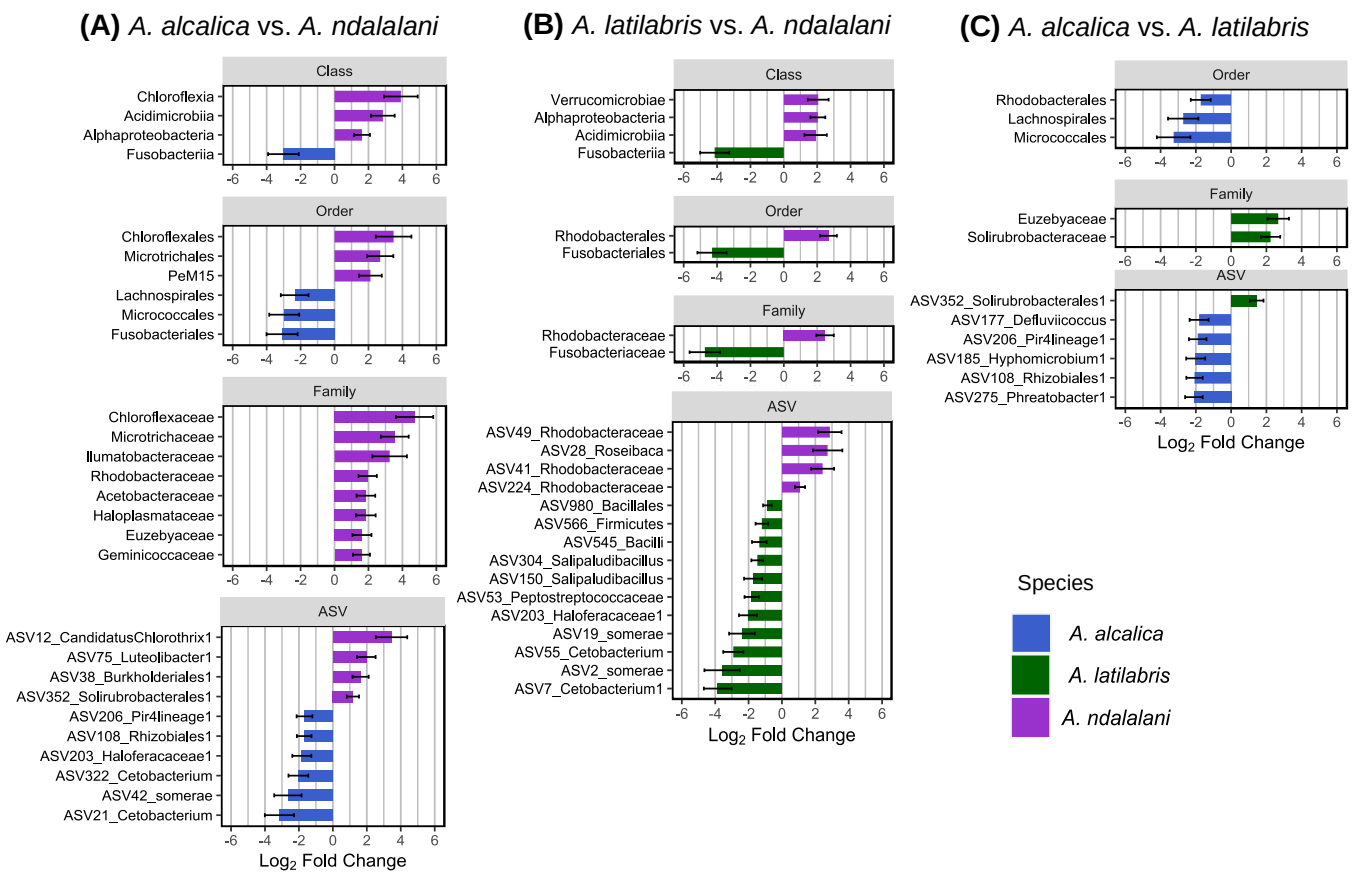


Figure A3.9. Differential abundance analysis of microbial taxa (16S rRNA) between sympatric Lake Natron *Alcolapia* species shows significantly enriched taxa between: (**A**) *A. alcalica* and *A. ndalalani,* (**B**) *A. latilabris* and *A. ndalalani,* (**C**) *A. alcalica* and *A. latilabris***.** Differentially abundant taxa between species were identified using LinDA (Zhou *et al.,* 2022) at 4 different phylogenetic levels (Class, Order, Family and ASVs). Only taxa that had a significant ($p<0.05$) Log2Fold Change are plotted.

Table A4.1. Sample information, collection coordinates and sequencing statistics for the whole genome re-sequenced *Alcolapia* samples.

Species	Sample ID	Site	Sampling GPS coordinates	Mapped reads/n	Mapped reads/%	Mean X coverage	Properly paired reads/%	Analyses	
								PCAngsd	*f*-branch statistics
A. alcalica (southern)	11A_322_AA_alcalicus_blue_CTTGTA	11A	S -2.591000 E 36.009445	28547289	98.10	5.80	0.88	✓	✓
	11A_323_AA_alcalicus_yellow_GTGAAA	11A	S -2.591000 E 36.009446	32859601	98.53	6.55	0.89	✓	✓
	11A_325_AA_alcalicus_yellow_ATGTCA	11A	S -2.591000 E 36.009447	34410207	98.99	6.73	0.89	✓	✓
A. alcalica (northern)	015_733_AU_alcalicus_upturn_GTGAAA	15	S -2.433361 E 36.10175	72254012	99.26	13.71	0.89	✓	✓
	015_734_AU_alcalicus_upturn_CTTGTA	15	S -2.433361 E 36.10175	67518677	97.82	12.94	0.88	✓	✓
	015_849_AA_alcalicus_blue_ATGTCA	15	S -2.433361 E 36.10175	62355047	98.77	11.94	0.88	✓	✓
A. latilabris	005_112_AL_latilabris_ACAGTG	5	S -2.597583 E 35.918417	72407847	98.88	13.70	0.88	✓	✓
	11B_419_AL_latilabris_GCCAAT	11B	S -2.591667 E 36.012056	66711548	99.14	12.58	0.88	✓	✓
A. ndalalani	11A_294_AN2_ndlalani_GCCAAT	11A	S -2.591000 E 36.009447	122617238	99.07	22.33	0.89	✓	✓
	005_127_AN2_ndlalani_ACAGTG	5	S -2.597583 E 35.918417	209019159	99.04	37.51	0.88	✓	✓
A. grahami	018_871_AG_grahami_magadi_ACAGTG	18	S -2.001139 E 36.231972	69583708	98.90	13.09	0.88	✓	✓
	021_899_AG_grahami_magadi_GCCAAT	21	S -1.844444 E 36.22425	16709109	99.24	3.64	0.90	✓	
	024_956_AG_grahami_Nakuru_CTTGTA	24	S -0.396028 E 36.107583	82919567	98.75	15.44	0.88	✓	✓
	024_969_AG_grahami_Nakuru_GTGAAA	24	S -0.396028 E 36.107583	15241323	98.98	3.38	0.88	✓	

Table A4.2. Priors and parameters used in the estimation file of the best fitting scenario for *fastsimcoal2*. Showing the search ranges (Minimum and Maximum) for each Simple parameter. The keywords "bounded" specifies that the upper range of the parameter is bounded, "paramInRange" specifies that T_DIV2 occurs after T_DIV1 in backwards coalescent time. Complex parameters are derived from simple parameters.

Type	Parameter	Minimum	Maximum	Keyword
Simple	N_AL	1000	1,000,000	bounded
	N_AA	1000	1,000,000	bounded
	N_AG	1000	1,000,000	bounded
	RS_AL	0.1	10	
	RS_AG_AA	0.1	10	
	RS_ANC	0.1	10	
	T_DIV1	1000	1,000,000	bounded
	T_DIV2	T_DIV1	10,000,000	paramInRange
	MIG_ANC1	1E-10	0.1	
	MIG_ANC2	1E-10	0.1	
	migr1_12	1E-10	0.1	
	migr1_21	1E-10	0.1	
	migr1_23	1E-10	0.1	
	migr1_32	1E-10	0.1	
Complex	N_AL_ANC	= N_AL * RS_AL		
	N_AG_AA	= N_AA * RS_AA		
	N_ANC	= N_AG_AA * RS_ANC		

Table A4.3. Estimated parameters from the best fitting model of all demographic scenarios tested. A-P refer to the topology and gene flow scenarios outlined in Figure 4.2.

Topology	Scenario	N_{AL}	N_{AA}	N_{AG}	T_{DIV1}	migr1_12	migr1_21	migr1_13	migr1_31	migr1_23	migr1_32	MIG_{ANC1}	MIG_{ANC2}	T_{DIV2}	N_{AL_ANC}	N_{AG_AA}	N_{ANC}	MaxEst Lhood	ΔlhoodΔ	AIC	ΔAIC
A. latilabris and *A. alcalica* as sister	A	69540	173339	110727	3196									15052	164816	385642	186855	-445143	197280.5	2049978	2312
	B	47125	183242	129727	4187	1.00E-04	3.80E-06							16947	251871	501796	138784	-445057	197194	2049584	1918
	C	59914	188200	57933	28883	1.20E-04	2.80E-05			8.30E-09	8.60E-05			156854	383261	1033849	19684	-444669	196806.6	2047802	136
	D	64449	170927	116384	2777							1.10E-07	4.40E-06	17408	222125	362479	150390	-445124	197261.5	2049895	2229
	E	85853	201375	74351	49204	7.20E-05	3.00E-05	9.40E-07	6.70E-07	4.20E-08	6.00E-05			268758	13962	771508	181242	-444709	196846.1	2047988	321
	F	92260	157557	94592	7086	3.90E-05	5.80E-05					3.90E-06	1.40E-07	19898	752983	760507	34990	-445071	197208.2	2049654	1987
	G	83077	192829	46261	24664	6.60E-05	4.00E-05			4.30E-08	1.00E-04	1.80E-09	1.40E-04	234406	256383	573410	14503	-444711	196848.1	2047999	333
	H	69760	175143	51668	23529	8.90E-05	3.90E-05	5.50E-09	8.90E-07	1.50E-08	1.00E-04	2.20E-08	1.40E-04	185381	18855	672930	340455	-444712	196849.3	2048009	343
A. grahami and *A. alcalica* as sister	I	124355	172211	45770	4255									7909	252041	1801406	293473	-445688	197825.3	2052488	4821
	J	18928	177071	109246	9379	4.00E-04	6.90E-07							208139	13092	523744	100717	-444675	196812.2	2047828	162
	K	17380	156476	72759	8691	4.60E-04	5.50E-08	7.20E-09	2.50E-05					198901	56509	630018	100416	-444660	196796.8	2047759	93
	L	37792	244398	90786	43775	1.80E-04	1.40E-05	1.40E-07	1.40E-06	2.80E-07	4.20E-05			767802	39804	105958	276790	-444713	196850.1	2048009	342
	M	63327	133798	41437	3446							2.40E-03	1.50E-02	301191	349622	197832	43754	-445145	197282.2	2049993	2326
	N	16098	171593	85998	8337	4.70E-04	5.80E-10					5.70E-05	3.50E-08	238465	54802	527298	67627	-444654	196791.7	2047738	71
	O	44619	168303	95879	11602	1.60E-04	2.60E-05	4.90E-09	2.40E-05			2.70E-03	1.70E-03	249438	141709	471230	67821	-444639	196775.8	2047666	-
	P	60867	189064	71925	33238	1.10E-04	3.40E-05	7.70E-08	2.90E-06	8.80E-08	6.50E-05	5.00E-03	2.70E-05	192634	85234	857227	96283	-444669	196806.6	2047814	148

REFERENCES

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J.L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J.Y., Prieto, P., Quesada, V. and Quilez, J. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology.*171(17), 1–19.

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R. and Butlin, R.K. (2013). Hybridization and speciation. *Journal of Evolutionary Biology* 26, 229–246.

Achenbach, L.A., Michaelidou, U., Bruce, R.A., Fryman, J. and Coates, J.D. (2001). Dechloromonas agitata gen. nov., sp. nov. and Dechlorosoma suillum gen. nov., sp. nov., two novel environmentally dominant (per) chlorate-reducing bacteria and their phylogenetic position. *International Journal of Systematic and Evolutionary Microbiology*, 51(2), 527-533.

Ackerly, D. D., Schwilk, D. W. and Webb, C. O. (2006). Niche evolution and adaptive radiation: testing the order of trait divergence. *Ecology*, 87(7), 50–61.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19, 716–723.

Albertson, R. C., Streelman, J. T., & Kocher, T. D. (2003). Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences*. 100(9), 5252-5257.

Albertson, R. C. (2008). Morphological Divergence Predicts Habitat Partitioning in a Lake Malawi Cichlid Species Complex. *Copeia* 2008, 689–698.

Albrecht H. (1968). Freiwasserbeobachtungen an Tilapien (Pisces, Cichlidae) in OstAfrika. Zeitschrift für Tierpsychologie, 25, 377–394. *Senckenberg Biologica*, 49, 107–118.

Albrecht H, Apfelbach R, Wickler W. (1968). Über die Eigenständigkeit der Art Tilapia grahami Boulenger, ihren Grubenbau und die Zucht in reines Susswasser (Pisces, Cichlidae). *Senckenberg Biologica*, 49, 107–118.

Almén, M. S., Lamichhaney, S., Berglund, J., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2016). Adaptive radiation of Darwin's finches revisited using whole genome sequencing. *BioEssays*, 38(1), 14-20.

Alekseyenko, A.V. (2016). Multivariate Welch t-test on distances. *Bioinformatics*, 32(23), 3552-3558.

Amillano-Cisneros, J.M., Hernández-Rosas, P.T., Gomez-Gil, B., Navarrete-Ramírez, P., Ríos-Durán, M.G., Martínez-Chávez, C.C., Johnston-Monje, D., Martínez-Palacios, C.A. and Raggi, L. (2022). Loss of gut microbial diversity in the cultured, agastric fish, Mexican pike silverside (Chirostoma estor: Atherinopsidae), *PeerJ*, 10, p. e13052.

Amoozegar, M.A., Shahinpei, A., Sepahy, A.A., Makhdoumi-Kakhki, A., Seyedmahdi, S.S., Schumann, P. and Ventosa, A. (2014). Pseudomonas salegens sp. nov., a halophilic member of the genus Pseudomonas isolated from a wetland, *International Journal of Systematic and Evolutionary Microbiology*, 64(10), 3565–3570.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arntzen, J. W., Jehle, R., Bardakci, F., Burke, T., and Wallis, G. P. (2009). Asymmetric viability of reciprocal-cross hybrids between crested and marbled newts (Triturus cristatus and T. marmoratus). *Evolution* 63, 1191–1202.

Astudillo-Clavijo, V., Stiassny, M. L. J., Ilves, K. L., Musilova, Z., Salzburger, W., and López-Fernández, H. (2023). Exon-based Phylogenomics and the Relationships of African Cichlid Fishes: Tackling the Challenges of Reconstructing Phylogenies with Repeated Rapid Radiations. *Systematic Biology*, 72(1), 134-149.

Babik, W., Butlin, R. K., Baker, W. J., Papadopulos, A. S., Boulesteix, M., Anstett, Lexer, C., Hutton, I.A.N. and Savolainen, V. (2009). How sympatric is speciation in the Howea palms of Lord Howe Island?. *Molecular Ecology,* 18(17), 3629-3638.

Baerends, G. P., & Baerends-van Roon, J. M. (1950). An introduction to the study of the ethology of the cichlid fishes. *Behaviour Supplement*. III-243.

Bakker, V. J., Finkelstein, M. E., D'elia, J., Doak, D. F., and Kirkland, S. (2022). Genetically based demographic reconstructions require careful consideration of generation time. *Current Biology* 32, 341–359.

Baldo, L., Santos, M. E., and Salzburger, W. (2011). Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome Biology and Evolution* 3, 443–455.

Baldo, L., Pretus, J.L., Riera, J.L., Musilova, Z., Bitja Nyom, A.R. and Salzburger, W., (2017). Convergence of gut microbiotas in the adaptive radiations of African cichlid fishes, *ISME Journal*, 11(9), 1975–1987.

Baldo, L., Riera, J. L., Salzburger, W., and Barluenga, M. (2019). Phylogeography and Ecological Niche Shape the Cichlid Fish Gut Microbiota in Central American and African Lakes. *Frontiers in Microbiology,* 10: 2372.

Banerjee, S., Azad, S. A., Vikineswary, S., Selvaraj, O. S., & Mukherjee, T. K. (2000). Phototrophic bacteria as fish feed supplement. *Asian-Australasian Journal of Animal Sciences*, 13(7), 991-994.

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*. 439(7077).

Barluenga, M., & Meyer, A. (2010). Phylogeography, colonization and population history of the Midas cichlid species complex (Amphilophus spp.) in the Nicaraguan crater lakes. *BMC Evolutionary Biology*, 10, 1-20.

Barraclough, T. G., and Nee, S. (2001). Phylogenetics and speciation. *Trends in Ecology and Evolution* 16, 391–399.

Barton, N. H. (1988). Speciation. *Analytical biogeography*, 185–218.

Basolo, A. L. (1990). Female Preference Predates the Evolution of the Sword in Swordtail Fish. *Science.* 250(4982), 808–810.

Bassil, N. M. and Lloyd, J. R. (2019). Anaerobacillus isosaccharinicus sp. Nov., an alkaliphilic bacterium which degrades isosaccharinic acid, *International Journal of Systematic and Evolutionary Microbiology*, 69(12), 3666–3671.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv*.1406. 5823.

Bayona, J. D. R. (2006). Alcolapia alcalicus. *IUCN Red List Threat. Species*. Available at: https://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T60452A12368336.en [Accessed February 13, 2023].

Bayona, J. D. R., and Akinyi, E. (2006). Alcolapia grahami. *IUCN Red List Threat. Species*. Available at: https://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T60453A12368415.en [Accessed February 13, 2023].

Beichman, A. C., Huerta-Sanchez, E., and Lohmueller, K. E. (2018). Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. *Annual Review of Ecology, Evolution, and Systematics* 49, 433–456.

Bergman, A. N., Laurent, P., Otiang'a-Owiti, G., Bergman, H. L., Walsh, P. J., Wilson, P., & Wood, C. M. (2003). Physiological adaptations of the gut in the Lake Magadi tilapia, Alcolapia grahami, an alkaline-and saline-adapted teleost fish. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 136(3), 701-715.

Berner, D., and Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics.* 31, 491–499.

Beule, L. and Karlovsky, P. (2020). Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. *PeerJ*, 8, p.e9593.

Beverton, R. J. H. (1992). Patterns of reproductive strategy parameters in some marine teleost fishes. *Journal of Fish Biology* 41, 137–160.

Bezault, E., Mwaiko, S., and Seehausen, O. (2011). Population genomic tests of models of adaptive radiation in Lake Victoria region cichlid fish. *Evolution,* 65, 3381–3397.

Bhute, S.S., Escobedo, B., Haider, M., Mekonen, Y., Ferrer, D., Hillyard, S.D., Friel, A.D., van Breukelen, F. and Hedlund, B.P. (2020). The gut microbiome and its potential role in paradoxical anaerobism in pupfishes of the Mojave Desert. *Animal Microbiome*, 2(1), 9.

Bird, C. E., Fernandez-Silva, I., Skillings, D. J., and Toonen, R. J. (2012). Sympatric Speciation in the Post "Modern Synthesis" Era of Evolutionary Biology. *Evolutionary Biology,* 39, 158–180.

Blais, J., Plenderleith, M., Rico, C., Taylor, M. I., Seehausen, O., van Oosterhout, C., & Turner, G. F. (2009). Assortative mating among Lake Malawi cichlid fish populations is not simply predictable from male nuptial colour. *BMC Evolutionary Biology*, 9, 1-12.

Bolnick, D. I., and Fitzpatrick, B. M. (2007). Sympatric Speciation: Models and Empirical Evidence. *Annual Reviews,* 38, 459–487.

Bolnick, D.I., Snowberg, L.K., Hirsch, P.E., Lauber, C.L., Knight, R., Caporaso, J.G. and Svanbäck, R. (2014). Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). *Ecology letters*, 17(8), 979-987.

Bolnick, D.I., Hund, A.K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. and Yukilevich, R. (2023). A multivariate view of the speciation continuum. *Evolution*, 77(1), 318-328.

Bootsma, H. A., Hecky, R. E., Hesslein, R. H., and Turner, G. F. (1996). Food Partitioning Among Lake Malawi Nearshore Fishes as Revealed by Stable Isotope Analyses. *Ecology,* 77, 1286–1290.

Borsodi, A.K., Aszalós, J.M., Bihari, P., Nagy, I., Schumann, P., Spröer, C., Kovács, A.L., Bóka, K., Dobosy, P., Óvári, M. and Szili-Kovács, T. (2019). Anaerobacillus alkaliphilus sp. Nov., a novel alkaliphilic and moderately halophilic bacterium, *International Journal of Systematic and Evolutionary Microbiology*, 69(3), 631–637.

Bosshard, P. P., Zbinden, R. and Altwegg, M. (2002). Turicibacter sanguinis gen. nov., sp. nov., a novel anaerobic, Gram-positive bacterium, *International Journal of Systematic and Evolutionary Microbiology*, 52(4), 1263–1266.

Boulenger, G. (1912). Description of a new fish from British East Africa. *Annals and Magazine of Natural History* 9(53), 519–521.

Bowman, J.P., Sly, L.I., Nichols, P.D. and Hayward, A.C. (1993). Revised taxonomy of the

methanotrophs: description of Methylobacter gen. nov., emendation of Methylococcus, validation of Methylosinus and Methylocystis species, and a proposal that the family Methylococcaceae includes only the group I methanotrophs. *International Journal of Systematic and Evolutionary Microbiology*, 43(4), 735-753.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E. and Turner-Maier, J. (2015). The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513, 375–381.

Breacker, C., Barber, I., Norton, W. H. J., McDearmid, J. R., & Tilley, C. A. (2017). A low-cost method of skin swabbing for the collection of DNA samples from small laboratory fish. *Zebrafish*. 14(1), 35-41.

Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J., & Bordenstein, S. R. (2016). Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biology*, 14(11), e2000225.

Brooks, M.E., Kristensen, K., Van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M. and Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2), 378-400.

Brown, W. L., & Wilson, E. O. (1956). Character displacement. *Systematic Zoology*, 5(2), 49-64.

Brucker, R. M. and Bordenstein, S. R. (2012). Speciation by symbiosis, *Trends in Ecology & Evolution*, 27(8), 443–451.

Burrough, S. L., and Thomas, D. S. G. (2009). Geomorphological contributions to palaeolimnology on the African continent. *Geomorphology,* 103, 285–298.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology,* 8(12), e1002822.

Butlin, R. (1987). Speciation by reinforcement. *Trends in Ecology & Evolution*, 2(1), 8-13.

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506), 2997-3007.

Butlin, R., Debelle, A., Kerth, C., Snook, R.R., Beukeboom, L.W., RF, C.C., Diao, W., Maan, M.E., Paolucci, S., Weissing, F.J. and van de Zande, L. (2012). What do we need to know about speciation? *Trends in Ecology & Evolution*. 27(1), 27-39.

Butzer, K. W., Isaac, G. L., Richardson, J. L., & Washbourn-Kamau, C. (1972). Radiocarbon Dating of East African Lake Levels: New observations provide fresh insights into late Quaternary paleoclimates. *Science*, 175(4026), 1069-1076.

Cai, L., Xi, Z., Lemmon, E.M., Lemmon, A.R., Mast, A., Buddenhagen, C.E., Liu, L. and Davis, C.C. (2021). The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology*, 70(3), 491-507.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581-583.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal*, 11(12), 2639-2643.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 1-9.

Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, 124(4), 561-583.

Cao, X., Liu, X. and Dong, X. (2003). Alkaliphilus crotonatoxidans sp. nov., a strictly anaerobic, crotonate-dismutating bacterium isolated from a methanogenic environment', *International Journal of Systematic and Evolutionary Microbiology*, 53(4), 971–975.

Casey, J. M., Meyer, C. P., Morat, F., Brandl, S. J., Planes, S., & Parravicini, V. (2019). Reconstructing hyperdiverse food webs: Gut content metabarcoding as a tool to disentangle trophic interactions on coral reefs. *Methods in Ecology and Evolution*, 10(8), 1157-1170.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* 1, 171–182.

Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., Collman, R.G., Bushman, F.D. and Li, H. (2012). Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, 28(16), 2106-2113.

Chen, J., King, E., Deek, R., Wei, Z., Yu, Y., Grill, D., & Ballman, K. (2018). An omnibus test for differential distribution analysis of microbiome sequencing data. *Bioinformatics*, 34(4), 643-651.

Clark, B., Elkin, J., Marconi, A., Turner, G.F., Smith, A.M., Joyce, D., Miska, E.A., Juntti, S.A. and Santos, M.E. (2022). Oca2 targeting using CRISPR/Cas9 in the Malawi cichlid Astatotilapia calliptera. *Royal Society Open Science*, 9(4), 220077.

Coe, Malcolm J. (1966). the biology of Tilapia grahami Boulenger in Lake Magadi, Kenya. *Acta Tropica*, 23, 146-177.

Coe, M. J. (1969). Observations on Tilapia alcalica Hilgendorf an endemic cichlid fish from Lake Natron, Tanzania. *Revue de Zoologie et de Botanique Africaines*, 801, 1-9.

Colombo, M., Indermaur, A., Meyer, B. S., & Salzburger, W. (2016). Habitat use and its implications to functional morphology: niche partitioning and the evolution of locomotory morphology in Lake Tanganyikan cichlids (Perciformes: Cichlidae). *Biological Journal of the Linnean Society*, 118(3), 536-550.

Conte, M. A., Gammerdinger, W. J., Bartie, K. L., Penman, D. J., & Kocher, T. D. (2017). A high quality assembly of the Nile Tilapia (Oreochromis niloticus) genome reveals the structure of two sex determination regions. *BMC genomics*, 18(1), 1-19.

Corral, P., Gutiérrez, M. C., Castillo, A. M., Domínguez, M., Lopalco, P., Corcelli, A., & Ventosa, A. (2013). Natronococcus roseus sp. nov., a haloalkaliphilic archaeon from a hypersaline lake. *International Journal of Systematic and Evolutionary Microbiology*, 63(1), 104-108.

Corral, P., Amoozegar, M. A., & Ventosa, A. (2019). Halophiles and their biomolecules: recent advances and future applications in biomedicine. *Marine Drugs*, 18(1), 33.

Corsaro, D., Thomas, V., Goy, G., Venditti, D., Radek, R., & Greub, G. (2007). 'Candidatus Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach Blatta orientalis (Insecta: Blattodea). *Systematic and Applied Microbiology*, 30(3), 221-228.

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can we name Earth's species before they go extinct?. *Science*, 339(6118), 413-416.

Coyne, Jerry A., & Orr, H. A. (1997). "Patterns of speciation in Drosophila" revisited. *Evolution*, 295-

303.

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sinauer Associates, Sunderland, MA.

Crozier, R. H. (1974). Niche shape and genetic aspects of character displacement. *American Zoologist*, 14(4), 1151-1157.

Cruickshank, T. E., and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology,* 23, 3133–3157.

Cruz, R., Carballo, M., Conde-Padín, P., & Rolán-Alvarez, E. (2004). Testing alternative models for sexual isolation in natural populations of Littorina saxatilis: Indirect support for by-product ecological speciation? *Journal of Evolutionary Biology* 17, 288–293.

Dahal, R. H., Chaudhary, D. K., Kim, D. U., & Kim, J. (2021). Luteolibacter luteus sp. nov., isolated from stream bank soil. *Archives of Microbiology*, 203, 377-382.

Dale, J., Dey, C. J., Delhey, K., Kempenaers, B., and Valcu, M. (2015). The effects of life history and sexual selection on male and female plumage colouration. *Nature.* 527, 367–370.

Darwin, C. & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society of London*. 3, 45–62

Darwin, C. (1859). *On the Origin of Species*. Routeledge.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, 6, 1-14.

Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L., Kartzinel, T.R. and Eveson, J.P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?. *Molecular Ecology*, 28(2), 391-406.

Debelius, J., Song, S. J., Vazquez-Baeza, Y., Xu, Z. Z., Gonzalez, A., & Knight, R. (2016). Tiny microbes, enormous impacts: what matters in gut microbiome studies?. *Genome Biology*, 17, 1-12.

Deino, A. L., Dommain, R., Keller, C. B., Potts, R., Behrensmeyer, A. K., Beverly, E. J., & Olorgesailie Drilling Project Scientific Team. (2019). Chronostratigraphic model of a high-resolution drill core record of the past million years from the Koora Basin, south Kenya Rift: Overcoming the difficulties of variable sedimentation rate and hiatuses. *Quaternary Science Reviews*, 215, 213-231.

Delamare-Deboutteville, J., Batstone, D.J., Kawasaki, M., Stegman, S., Salini, M., Tabrett, S., Smullen, R., Barnes, A.C. and Hülsen, T. (2019). Mixed culture purple phototrophic bacteria is an effective fishmeal replacement in aquaculture. *Water research* X, 4, 100031.

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*. 400(6742), 354-357.

Dieckmann, U., Doebeli, M., Metz, J. A. J., and Tautz, D. (2012). *Adaptive speciation*. Cambridge University Press.

Dill-McFarland, K. A., Weimer, P. J., Pauli, J. N., Peery, M. Z., & Suen, G. (2016). Diet specialization selects for an unusual and simplified gut microbiota in two-and three-toed sloths. *Environmental Microbiology*, 18(5), 1391-1402.

Djemiel, C., Plassard, D., Terrat, S., Crouzet, O., Sauze, J., Mondy, S., Nowak, V., Wingate, L., Ogée, J. and Maron, P.A. (2020). µgreen-db: a reference database for the 23S rRNA gene of eukaryotic plastids and cyanobacteria. *Scientific Reports*, 10(1), 5915.

Dobzhansky, T. (1937). Genetic nature of species differences. *The American Naturalist*, 71(735), 404-420.

Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *The American Naturalist*, 74(753), 312-321.

Dobzhansky, T. (1982). Genetics and the Origin of Species (No. 11). Columbia university press.

Dommain, R., Riedl, S., Olaka, L.A., deMenocal, P., Deino, A.L., Owen, R.B., Muiruri, V., Müller, J., Potts, R. and Strecker, M.R. (2022). Holocene bidirectional river system along the Kenya Rift and its influence on East African faunal exchange and diversity gradients. *Proceedings of the National Academy of Sciences*, 119(28), e2121388119.

Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* 28, 2239–2252.

Drès, M., & Mallet, J. (2002). Host races in plant–feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1420), 471-492.

Dvořák, P., Hindák, F., Hašler, P., Hindakova, A., and Poulíčková, A. (2014). Morphological and molecular studies of Neosynechococcus sphagnicola, gen. et sp. nov.(Cyanobacteria, Synechococcales). *Phytotaxa*, 170(1), 24-34.Edwin, N., Akhwale, J. K. and Mwirichia, R. (2019). Phylogenetic and morphological diversity of culturable cyanobacteria from Lake Magadi in Kenya, *African Journal of Biological Sciences (South Africa),* 1(4), 24–31.

Egger, B., Mattersdorfer, K., and Sefc, K. M. (2010). Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. *Journal of Evolutionary Biology* 23, 433–439.

Elmer, K. R., Lehtonen, T. K., & Meyer, A. (2009). Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. *Evolution*. 63(10), 2750-2757.

Elmer, K. R., Kusche, H., Lehtonen, T. K., & Meyer, A. (2010). Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1547), 1763-1782.

Enciso-Romero, J., Pardo-Díaz, C., Martin, S.H., Arias, C.F., Linares, M., McMillan, W.O., Jiggins, C.D. and Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, 26(19), pp.5160-5172.

Eschmeyer, W. N., Fricke, R., and Van der Laan, R. (2022). Species in the genus Oreochromis. *Catalog of Fishes*. California Academy of Sciences. Retrieved 20 January 2022.

Eugster, H. P. (1986). Lake Magadi, Kenya: a model for rift valley hydrochemistry and sedimentation? *Geological Society, London, Special Publications*, 25(1), 177-189.

Evans, K. L., Warren, P. H. and Gaston, K. J. (2005) Species–energy relationships at the macroecological scale: a review of the mechanisms, *Biological Reviews*, 80(1), 1–25.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., and Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLOS Genetics* 9, e1003905.

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., and Sousa, V. C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* 37, 4882–4885.

Faria, R., Renaut, S., Galindo, J., Pinho, C., Melo-Ferreira, J., Melo, M., Jones, F., Salzburger, W., Schluter, D. and Butlin, R. (2014). Advances in Ecological Speciation: an integrative approach. *Molecular Ecology*, 23, 513-521.

Farmer Iii, J. J., Michael Janda, J., Brenner, F. W., Cameron, D. N., & Birkhead, K. M. (2015). Vibrio. *Bergey's Manual of Systematics of Archaea and Bacteria.* 1-79.

Feder, J. L., Chilcote, C. A., and Bush, G. L. (1988). Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella. *Nature* 336, 61–64.

Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J., and Romero-Severson, J. (2003). Evidence for Inversion Polymorphism Related to Sympatric Host Race Formation in the Apple Maggot Fly, Rhagoletis pomonella. *Genetics* 163, 939–953.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342-350.

Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J., & Romero-Severson, J. (2003). Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, Rhagoletis pomonella. *Genetics*, 163(3), 939-953.

Fendrich, C. (1988). Halovibrio variabilis gen. nov. sp. nov., Pseudomonas halophila sp. nov. and a New Halophilic Aerobic Coccoid Eubacterium from Great Salt Lake, Utah, USA, *Systematic and Applied Microbiology*, 11(1), pp. 36–43.

Fendrich, C., Hippe, H. and Gottschalk, G. (1990). Clostridium halophilium sp. nov. and C. litorale sp. nov., an obligate halophilic and a marine species degrading betaine in the Stickland reaction, *Archives of Microbiology,* 154(2), pp. 127–132.

Fisher, R. A. (1958). The genetical theory of natural selection. Clarendon Press.

Fitzpatrick, B. M., Fordyce, J. A., and Gavrilets, S. (2008). What, if anything, is sympatric speciation? *Journal of Evolutionary Biology* 21, 1452–1459.

Fitzpatrick, S.W., Gerberich, J.C., Angeloni, L.M., Bailey, L.L., Broder, E.D., Torres-Dowdall, J., Handelsman, C.A., López-Sepulcre, A., Reznick, D.N., Ghalambor, C.K. and Chris Funk, W. (2016). Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. *Evolutionary Applications*, 9(7), 879-891.

Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology,* 28(3), 544-567.

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2020). A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, 37(4), 1211-1223.

Foflonker, F., Ananyev, G., Qiu, H., Morrison, A., Palenik, B., Dismukes, G. C., & Bhattacharya, D. (2016). The unexpected extremophile: Tolerance to fluctuating salinity in the green alga. *Algal Research,* 16, 465-472.

Foote, A. D. (2018). Sympatric Speciation in the Genomic Era. *Trends in Ecology and Evolution* 33, 85–95.

Ford, A. G., Dasmahapatra, K. K., Rüber, L., Gharbi, K., Cezard, T., & Day, J. J. (2015). High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*, 24(13), 3421-3440.

Ford, A. G. P., Rüber, L., Newton, J., Dasmahapatra, K. K., Balarin, J. D., Bruun, K., & Day, J. J. (2016). Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution,* 70(12), 2718-2735.

Ford, A.G., Bullen, T.R., Pang, L., Genner, M.J., Bills, R., Flouri, T., Ngatunga, B.P., Rüber, L., Schliewen, U.K., Seehausen, O. and Shechonge, A. (2019). Molecular phylogeny of Oreochromis (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple

colonisation of adverse aquatic environments. *Molecular Phylogenetics and Evolution*, 136, 215-226.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S. and Heiberger, R. (2012). Package 'car'. Vienna: R Foundation for Statistical Computing, 16.

Franchini, P., Fruciano, C., Frickey, T., Jones, J. C., & Meyer, A. (2014). The gut microbial community of Midas cichlid fish in repeatedly evolved limnetic-benthic species pairs. *PloS One*, 9(4), e95027.

Frantz, L.A., Schraiber, J.G., Madsen, O., Megens, H.J., Bosse, M., Paudel, Y., Semiadi, G., Meijaard, E., Li, N., Crooijmans, R.P. and Archibald, A.L. (2013). Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biology*, 14, 1-12.

Funk, D. J. (1998). Isolating a role for natural selection in speciation: host adaptation and sexual isolation in Neochlamisus bebbianae leaf beetles. *Evolution*. 52(6), 1744-1759.

Galvez, J. R., St. John, M. E., McLean, K., Dening Touokong, C., Gonwouo, L. N., & Martin, C. H. (2022). Trophic Specialization on unique resources despite limited niche divergence in a celebrated example of sympatric speciation. *Ecology of Freshwater Fish*, 31(4), 675-692.

Gavrilets, S. (2014). Is sexual conflict an "engine of speciation"?. *Cold Spring Harbor Perspectives in Biology*, 6(12), a017723.

Gavrilets, S., & Losos, J. B. (2009). Adaptive radiation: contrasting theory with data. *Science*, 323(5915), 732-737.

Genner, M. J., Turner, G. F., Barker, S., & Hawkins, S. J. (1999a). Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecology Letters*, 2(3), 185-190.

Genner, M. J., Turner, G. F., & Hawkins, S. J. (1999b). Foraging of rocky habitat cichlid fishes in Lake Malawi: coexistence through niche partitioning?. *Oecologia*, 121, 283-292.

Genner, M. J., & Turner, G. F. (2005). The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries*, 6(1), 1-34.

Genner, M. J., Nichols, P., Carvalho, G. R., Robinson, R. L., Shaw, P. W., Smith, A., & Turner, G. F. (2007). Evolution of a cichlid fish in a Lake Malawi satellite lake. *Proceedings of the Royal Society B: Biological Sciences,* 274(1623), 2249-2257.

Genner, M. J., Young, K. A., Haesler, M. P., and Joyce, D. A. (2008). Indirect mate choice, direct mate choice and species recognition in a bower-building cichlid fish lek. *Journal of Evolutionary Biology* 21, 1387–1396.

Genner, M. J., & Turner, G. F. (2012). Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution*, 29(1), 195-206.

Ghanbari, M., Kneifel, W. and Domig, K. J. (2015). A new view of the fish gut microbiome: Advances from next-generation sequencing, *Aquaculture*, 448, 464–475.

Giatsis, C., Sipkema, D., Smidt, H., Heilig, H., Benvenuti, G., Verreth, J., & Verdegem, M. (2015). The impact of rearing environment on the development of gut microbiota in tilapia larvae. *Scientific Reports,* 5(1), 1-15.

Girard, L., Lood, C., Höfte, M., Vandamme, P., Rokni-Zadeh, H., van Noort, V., Lavigne, R. and De Mot, R. (2021). The ever-expanding Pseudomonas genus: description of 43 new species and partition of the Pseudomonas putida group. *Microorganisms*, 9(8), 1766.

Givnish, T. J. (2015). Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytologist,* 207, 297–303.

Goodwin, S. and Zeikus, J. G. (1987). Physiological adaptations of anaerobic bacteria to low pH: metabolic control of proton motive force in Sarcina ventriculi, *Journal of Bacteriology*, 169(5), 2150–2157.

Gould, A.L., Zhang, V., Lamberti, L., Jones, E.W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J.M., Beerenwinkel, N. and Ludington, W.B. (2018). Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences*, 115(51), E11951-E11960.

Grant, P. R., & Grant, B. R. (1997). Genetics and the origin of bird species. *Proceedings of the National Academy of Sciences*. 94(15), 7768-7775.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y. and Hansen, N.F. (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979), 710-722.

Greene, L.K., Williams, C.V., Junge, R.E., Mahefarisoa, K.L., Rajaonarivelo, T., Rakotondrainibe, H., O’Connell, T.M. and Drea, C.M. (2020). A role for gut microbiota in host niche differentiation. *The ISME Journal*, 14(7), 1675-1687.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., and Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics.* 43, 1031–1035.

Groussin, M., Mazel, F., and Alm, E. J. (2020). Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host & Microbe* 28, 12–22.

Guerrero, R. F., and Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 26, 5362–5368.

Guilloux, C. A., Lamoureux, C., Beauruelle, C., & Héry-Arnaud, G. (2021). Porphyromonas: A neglected potential key genus in human microbiomes. *Anaerobe*, 68, 102230.

Gupta, R. S., Lo, B., & Son, J. (2018). Phylogenomics and comparative genomic studies robustly support division of the genus Mycobacterium into an emended genus Mycobacterium and four novel genera. *Frontiers in Microbiology*, 9, 67.

Habe, H., Kanemitsu, M., Nomura, M., Takemura, T., Iwata, K., Nojiri, H., Yamane, H. and Omori, T. (2004). Isolation and characterization of an alkaliphilic bacterium utilizing pyrene as a carbon source. *Journal of Bioscience and Bioengineering*, 98(4), 306-308.

Hajibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLoS One*, 14(9), e0220953.

Hamidi, B., Wallace, K., Vasu, C., & Alekseyenko, A. V. (2019). Wd∗-test: robust distance-based multivariate analysis of variance. *Microbiome*, 7(1), 1-9.

Hänfling, B., and Weetman, D. (2006). Concordant Genetic Estimators of Migration Reveal Anthropogenically Enhanced Source-Sink Population Structure in the River Sculpin, Cottus gobio. *Genetics* 173, 1487–1501.

Härer, A., Torres-Dowdall, J., Rometsch, S. J., Yohannes, E., Machado-Schiaffino, G., and Meyer, A. (2020). Parallel and non-parallel changes of the gut microbiota during trophic diversification in repeated young adaptive radiations of sympatric cichlid fish. *Microbiome* 8, 1–14.

Harrison, A. D. (2002). A general view on the communities of the chironomidae adults in the Ethiopian Rift Valley lakes. Ethiopian Rift Valley Lakes (Tudorancea, C. & Taylor, WD, eds), 157-162.

Hartig, F. (2020). DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. *R package.*

Hata, H., Tanabe, A. S., Yamamoto, S., Toju, H., Kohda, M., & Hori, M. (2014). Diet disparity among sympatric herbivorous cichlids in the same ecomorphs in Lake Tanganyika: amplicon pyrosequences on algal farms and stomach contents. *BMC Biology*, 12, 1-14.

Heathcote, R.J., While, G.M., MacGregor, H.E., Sciberras, J., Leroy, C., D'Ettorre, P. and Uller, T. (2016). Male behaviour drives assortative reproduction during the initial stage of secondary contact. *Journal of Evolutionary Biology*, 29(5), 1003-1015.

Hendry, A. P. (2009). Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences,* 66, 1383–1398.

Henning, F., & Meyer, A. (2014). The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. *Annual Review of Genomics and Human Genetics*, 15, 417-441.

Henry, L. P., Bruijning, M., Forsberg, S. K., & Ayroles, J. F. (2021). The microbiome extends host evolutionary potential. *Nature Communications*, 12(1), 5141.

Henshaw, J. M., Kahn, A. T., & Fritzsche, K. (2016). A rigorous comparison of sexual selection indexes via simulations of diverse mating systems. *Proceedings of the National Academy of Sciences*, 113(3), E300-E308.

Hilgendorf, F. M. (1905). Fische von Deutsch und Englisch Ost-Africa. Zoologische Jahrbücher. Abteilung für Systematik. *Geographie und Biologie der Tiere (Jena)*, 22, 405-420.

Hillaire-Marcel, C., Carro, O., and Casanova, J. (1986). 14C and Th/U dating of Pleistocene and Holocene stromatolites from East African paleolakes. *Quaternary Research,* 25, 239–312.

Horn, M. A., Ihssen, J., Matthies, C., Schramm, A., Acker, G., & Drake, H. L. (2005). Dechloromonas denitrificans sp. nov., Flavobacterium denitrificans sp. nov., Paenibacillus anaericanus sp. nov. and Paenibacillus terrae strain MH72, N2O-producing bacteria isolated from the gut of the earthworm Aporrectodea caliginosa. *International Journal of Systematic and Evolutionary Microbiology*, 55(3), 1255-1265.

Houck, L. D., & Verrell, P. A. (1993). Studies of courtship behavior in plethodontid salamanders: a review. *Herpetologica*, 49(2), 175-184.

Huber, S. K., León, L. F. D., Hendry, A. P., Bermingham, E., & Podos, J. (2007). Reproductive isolation of sympatric morphs in a population of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, 274(1619), 1709-1714.

Hunter, M. I. S., Olawoye, T. L., & Saynor, D. A. (1981). The effect of temperature on the growth and lipid composition of the extremely halophilic coccus, Sarcina marina. *Antonie van Leeuwenhoek*, 47, 25-40.

Hulsey, C. D. (2005). Function of a key morphological innovation: fusion of the cichlid pharyngeal jaw. *Proceedings of the Royal Society B: Biological Sciences,* 273(1587), 669-675.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J. and Tappu, R., (2016). MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, 12(6), p.e1004957.

Igeño, M. I., Del Moral, C. G., Castillo, F., & Caballero, F. J. (1995). Halotolerance of the phototrophic bacterium Rhodobacter capsulatus E1F1 is dependent on the nitrogen source. *Applied and Environmental Microbiology*, 61(8), 2970-2975.

Imhoff, J. F. (2015). Rhodobacter, *Bergey's Manual of Systematics of Archaea and Bacteria*, 1–12. Hoboken, NJ.

Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P., Fischer, C., Lemmon, A.R., Lemmon, E.M., Thallinger, G.G. and Sturmbauer, C. (2018). Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications*, 9(1), 3159.

Irwin, D., & Schluter, D. (2022). Hybridization and the Coexistence of Species. *The American Naturalist*, 200(3), E93-E109.

Itoi, S., Abe, T., Washio, S., Ikuno, E., Kanomata, Y., & Sugita, H. (2008). Isolation of halotolerant Lactococcus lactis subsp. lactis from intestinal tract of coastal fish, *International Journal of Food Microbiology*, 121(1), 116–121.

Ivory, S. J., Blome, M. W., King, J. W., McGlue, M. M., Cole, J. E., & Cohen, A. S. (2016). Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proceedings of the National Academy of Sciences,* 113(42), 11895-11900.

Jakubavičiūtė, E., Bergström, U., Eklöf, J. S., Haenel, Q., & Bourlat, S. J. (2017). DNA metabarcoding reveals diverse diet of the three-spined stickleback in a coastal ecosystem. *PLoS One*, 12(10), e0186929.

Jamie, G. A., and Meier, J. I. (2020). The Persistence of Polymorphisms across Species Radiations. *Trends in Ecology and Evolution* 35, 795–808.

Jiang, F., Li, W., Xiao, M., Dai, J., Kan, W., Chen, L., Li, W., Fang, C. and Peng, F. (2012). Luteolibacter luojiensis sp. nov., isolated from Arctic tundra soil, and emended description of the genus Luteolibacter. *International Journal of Systematic and Evolutionary Microbiology*, 62(9), 2259-2263.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S. and Soltis, D.E. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473(7345), 97-100.

Johannsson, O.E., Bergman, H.L., Wood, C.M., Laurent, P., Kavembe, D.G., Bianchini, A., Maina, J.N., Chevalier, C., Bianchini, L.F., Papah, M.B. and Ojoo, R.O. (2014). Air breathing in Magadi tilapia Alcolapia grahami, under normoxic and hyperoxic conditions, and the association with sunlight and reactive oxygen species. *Journal of Fish Biology*, 84(3), 844-863.

Jonasson, J., Harkonen, T., Sundqvist, L., Edwards, S. V., & Harding, K. C. (2022). A unifying framework for estimating generation time in age-structured populations: implications for phylogenetics and conservation biology. *The American Naturalist*, 200(1), 48-62.

Jones, B. F., Eugster, H. P., & Rettig, S. L. (1977). Hydrochemistry of the lake Magadi basin, Kenya. *Geochimica et Cosmochimica Acta*, 41(1), 53-72.

Jones, B. E., & Grant, W. D. (1999). Microbial diversity and ecology of the Soda Lakes of East Africa. *In Microbial Biosystems: New Frontiers: Proceedings of the 8th International Symposium for Microbial Ecology (2000). Atlantic Canada Society for Microbial Ecology*, (681-687).

Jones, O. R., & Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10(3), 551-555.

Jordan, R., Kellogg, K., Juanes, F., & Stauffer, J., Jr. (2003). Evaluation of female mate choice cues in a group of Lake Malawi mbuna (Cichlidae). *Copeia,* (1), 181-186.

Joshi, N. A., & Fass, J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.

Joyce, D. A., Lunt, D. H., Genner, M. J., Turner, G. F., Bills, R., & Seehausen, O. (2011). Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology*, 21(3), R108-R109.

Kadigi, R. M., Mwathe, K., Dutton, A., Kashaigili, J., & Kilima, F. (2014). Soda ash mining in lake Natron: a reap or ruin for Tanzania?. *Journal of Environmental Conservation Research,* 2(2), 37-49.

Kalacska, M., Arroyo-Mora, J. P., Lucanus, O., & Kishe-Machumu, M. A. (2017). Land cover, land use, and climate change impacts on endemic cichlid habitats in northern Tanzania. *Remote Sensing*, 9(6), 623.

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16(5), 1099-1106.

Kambura, A. K., Mwirichia, R. K., Kasili, R. W., Karanja, E. N., Makonde, H. M., & Boga, H. I. (2016). Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. *BMC Microbiology*, 16(1), 1-12.

Kanai, H., Kobayashi, T., Aono, R., & Kudo, T. (1995). Natronococcus amylolyticus sp. nov., a haloalkaliphilic archaeon. *International Journal of Systematic Bacteriology*, 45(4), 762-766.

Kashinskaya, E. N., Simonov, E. P., Kabilov, M. R., Izvekova, G. I., Andree, K. B., & Solovyev, M. M. (2018). Diet and other environmental factors shape the bacterial communities of fish gut in an eutrophic lake. *Journal of Applied Microbiology*, 125(6), 1626-1641.

Kaufman, A., Margaritz, M., Paul, M., Hillaire-Marcel, C., Hollos, G., Boaretto, E., & Taieb, M. (1990). The 36C1 ages of the brines in the Magadi-Natron basin, East Africa. *Geochimica et Cosmochimica Acta*, 54(10), 2827-2833.

Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2016). Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genetics*, 12(6), e1006157.

Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2018). Lessons from a natural experiment: Allopatric morphological divergence and sympatric diversification in the Midas cichlid species complex are largely influenced by ecology in a deterministic way. *Evolution Letters*, 2(4), 323-340.

Kautt, A.F., Kratochwil, C.F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Hulsey, C.D., Franchini, P. and Pippel, M., (2020). Contrasting signatures of genomic divergence during sympatric speciation. *Nature*, 588(7836), 106-111.

Kavembe, G. D., Machado-Schiaffino, G., & Meyer, A. (2014). Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. *Hydrobiologia*, 739(1), 55-71.

Kavembe, G. D., Franchini, P., Irisarri, I., Machado-Schiaffino, G., & Meyer, A. (2015). Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one of earth's most extreme environments, the hypersaline soda Lake Magadi in Kenya, east Africa. *Journal of Molecular Evolution*, 81, 90-109.

Kavembe, G. D., Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2016). Eco-morphological differentiation in Lake Magadi tilapia, an extremophile cichlid fish living in hot, alkaline and hypersaline lakes in East Africa. *Molecular Ecology,* 25(7), 1610-1625.

Keller-Costa, T., Canário, A. V., & Hubbard, P. C. (2015). Chemical communication in cichlids: a mini-review. *General and Comparative Endocrinology*, 221, 64-74.

Keller, I., Wagner, C.E., Greuter, L., Mwaiko, S., Selz, O.M., Sivasundar, A., Wittwer, S. and Seehausen, O., (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, 22(11), 2848-2863.

Kevrekidis, C., Ruthensteiner, B., Cerwenka, A. F., Penk, S. B., & Reichenbacher, B. (2020). New cichlid fossils from the middle-late Miocene alkaline lakes of Africa. *Journal of Vertebrate Paleontology,* 40(4), e1805621.

Kim, S., Kang, I., & Cho, J. C. (2017). Genomic Analysis of a Freshwater Actinobacterium. *Journal of Microbiology and Biotechnology*, 27(4), 825-833.

Kim, P.S., Shin, N.R., Lee, J.B., Kim, M.S., Whon, T.W., Hyun, D.W., Yun, J.H., Jung, M.J., Kim, J.Y. and Bae, J.W. (2021). Host habitat is the major determinant of the gut microbiome of fish. *Microbiome*, 9(1), 166.

Kipkemboi, J. (2016). Vascular plants in eastern Africa rift valley saline wetlands. *Soda Lakes East Africa*, 285–293.

Klappenbach, J. A., & Pierson, B. K. (2004). Phylogenetic and physiological characterization of a filamentous anoxygenic photoautotrophic bacterium 'Candidatus Chlorothrix halophila'gen. nov., sp. nov., recovered from hypersaline microbial mats. *Archives of Microbiology*, 181, 17-25.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1.

Knight, M. E., Turner, G. F., Rico, C., Van Oppen, M. J. H., & Hewitt, G. M. (1998). Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology,* 7(11), 1605-1610.

Knight, M. E., & Turner, G. F. (1999). Reproductive isolation among closely related Lake Malawi cichlids: can males recognize conspecific females by visual cues?. *Animal Behaviour*, 58(4), 761-768.

Knight, M. E., & Turner, G. F. (2004). Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish Pseudotropheus zebra from Lake Malawi. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1540), 675-680.

Kocher, T. D. (2004). Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, 5(4), 288-298.

Kofler, R., Schlötterer, C., & Lelley, T. (2007). SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, 23(13), 1683-1685.

Kohl, K. D. (2020). Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. *Philosophical Transactions of the Royal Society B*, 375(1798), 20190251.

Komárek, J., & Johansen, J. R. (2015). Filamentous cyanobacteria. In *Freshwater Algae of North America,* (135-235). Academic Press.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 1-13.

Kostanjšek, R., Štrus, J., Drobne, D., & Avguštin, G. (2004). 'Candidatus Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod Porcellio scaber (Crustacea: Isopoda). *International Journal of Systematic and Evolutionary Microbiology*, 54(2), 543-549.

Kraaijeveld, K., Kraaijeveld-Smit, F. J., & Maan, M. E. (2011). Sexual selection and speciation: the comparative evidence revisited. *Biological Reviews*, 86(2), 367-377.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, 464(7290), 894-897.

Krienitz, L., Krienitz, D., Dadheech, P.K., Hübener, T., Kotut, K., Luo, W., Teubner, K. and Versfeld, W.D. (2016). Food algae for Lesser Flamingos: a stocktaking. *Hydrobiologia,* 775, 21–50.

Krienitz, L., and Schagerl, M. (2016). Tiny and tough: Microphytes of East African Soda Lakes. *Soda Lakes of East Africa*, 149–177.

Krienitz, L., Bock, C., Kotut, K., & Luo, W. (2012). Picocystis salinarum (Chlorophyta) in saline lakes and hot springs of East Africa. *Phycologia*, 51(1), 22-32.

Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., & Westram, A. M. (2020). Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B*, 375(1806), 20190528.

Kutschera, V. E., Bidon, T., Hailer, F., Rodi, J. L., Fain, S. R., & Janke, A. (2014). Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Molecular Biology and Evolution*, 31(8), 2004-2017.

Labrenz, M., Lawson, P. A., Tindall, B. J., & Hirsch, P. (2009). Roseibaca ekhonensis gen. nov., sp. nov., an alkalitolerant and aerobic bacteriochlorophyll a-producing alphaproteobacterium from hypersaline Ekho Lake. *International Journal of Systematic and Evolutionary Microbiology*, 59(8), 1935-1940.

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular ecology*, 28(2), 420-430.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78(6), 3721-3725.

Lande, R. and Kirkpatrick, M. (1988). Ecological speciation by sexual selection, *Journal of Theoretical Biology*, 133(1), pp. 85–98.

Lawson, P. A. and Rainey, F. A. (2016). Proposal to restrict the genus Clostridium prazmowski to Clostridium butyricum and related species, *International Journal of Systematic and Evolutionary Microbiology*, 66(2), 1009–1016.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Estimated marginal means, aka least-squares means. *R Package*.

Lemaire, O. N., Méjean, V., & Iobbi-Nivol, C. (2020). The Shewanella genus: ubiquitous organisms sustaining and preserving aquatic ecosystems. *FEMS Microbiology Reviews*, 44(2), 155-170.

Leray, M., Meyer, C. P., & Mills, S. C. (2015). Metabarcoding dietary analysis of coral dwelling predatory fish demonstrates the minor contribution of coral mutualists to their highly partitioned, generalist diet. *PeerJ*, 3, e1047.

Lewanski, A. L., Golcher-Benavides, J., Rick, J. A., & Wagner, C. E. (2022). Variable hybridization between two Lake Tanganyikan cichlid species in recent secondary contact. *Molecular Ecology*, 31(19), 5041-5059.

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., & Gordon, J. I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, 6(10), 776-788.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Li, H. (2011). Tabix: Fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* 27, 718–719.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* preprint arXiv:1303.3997.

Liem, K. F. (1973). Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Systematic Zoology*, 22(4), 425-441.

Liem, K. F. (1980) Adaptive Significance of Intra- and Interspecific Differences in the Feeding Repertoires of Cichlid Fishes, *Integrative and Comparative Biology*, 20(1), 295–314.

Logan, N. A. and Vos, P.D. (2015) Bacillus' *Bergey's Manual of Systematics of Archaea and Bacteria*, 1–163.

Loo, W. T., García-Loor, J., Dudaniec, R. Y., Kleindorfer, S., & Cavanaugh, C. M. (2019). Host phylogeny, diet, and habitat differentiate the gut microbiomes of Darwin's finches on Santa Cruz Island. *Scientific Reports*, 9(1), 18781.

Losos, J. B. (2011). *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles* (Vol. 10). University of California Press.

Pullin, R. S., & Lowe-McConnell, R. H. (1982). The Biology and Culture of Tilapias: Proceedings of the International Conference on the Biology and Culture of Tilapias, 2-5 September 1980 at the Study and Conference Center of the Rockefeller Foundation, Bellagio, Italy (Vol. 7). WorldFish.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4(12), 981-994.

Maan, M. E., Hofker, K. D., van Alphen, J. J., & Seehausen, O. (2006). Sensory drive in cichlid speciation. *The American Naturalist*, 167(6), 947-954.

Maan, M. E., & Seehausen, O. (2010). Mechanisms of species divergence through visual adaptation and sexual selection: perspectives from a cichlid model system. *Current Zoology*, 56(3), 285-299.

Maan, M. E., Seehausen, O. L. E., & Van Alphen, J. J. (2010). Female mating preferences and male coloration covary with water transparency in a Lake Victoria cichlid fish. *Biological Journal of the Linnean Society*, 99(2), 398-406.

Maan, M. E., and Seehausen, O. (2011). Ecology, sexual selection and speciation. *Ecology Letters,* 14, 591–602.

Maan, M. E., and Sefc, K. M. (2013). Colour variation in cichlid fish: Developmental mechanisms , selective pressures and evolutionary consequences. *Seminars in Cell and Developmental Biology,* 24, 516–528.

Mable, B. K., Alexandrou, M. A., & Taylor, M. I. (2011). Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology*, 284(3), 151-182.

Machado-Schiaffino, G., Henning, F., and Meyer, A. (2014). Species-specific differences in adaptive phenotypic plasticity in an ecologically relevant trophic trait: Hypertrophic lips in midas cichlid fishes. *Evolution,* 68(7), 2086–2091.

Machado-Schiaffino, G., Kautt, A.F., Torres-Dowdall, J., Baumgarten, L., Henning, F. and Meyer, A. (2017). Incipient speciation driven by hypertrophied lips in Midas cichlid fishes? *Molecular Ecology,* 26(8), 2348-2362.

Madigan, M. T., Schaaf, N. A. V., & Sattley, W. M. (2017). The chlorobiaceae, chloroflexaceae, and heliobacteriaceae. In: *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects*, 139-161. Springer.

Maki, J. J. and Looft, T. (2022) Turicibacter bilis sp. nov., a novel bacterium isolated from the chicken eggshell and swine ileum, *International Journal of Systematic and Evolutionary Microbiology*, 72(1), 005153.

Maina, J.N., Kavembe, G.D., Papah, M.B., Mashiteng, R., Wood, C.M., Bianchini, A., Bianchini, L.F., Bergman, H.L., Johannsson, O.E., Laurent, P. and Chevalier, C. (2019). Sizes, condition factors and sex ratios of the scattered populations of the small cichlid fish, Alcolapia grahami, that inhabits the lagoons and sites of Lake Magadi (Kenya), one of the most extreme aquatic habitats on Earth. *Environmental Biology of Fishes*, 102, 1265-1280.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J. and Turner, G.F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350(6267), 1493-1498.

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018a). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2(12), 1940-1955.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018b). RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution*, 35(5), 1284-1290.

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584-595.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology and Evolution,* 10, 294–299.

Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279-283.

Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363, 2971–2986.

Mallet, J., Meyer, A., Nosil, P., and Feder, J. L. (2009). Space, sympatry and speciation. *Journal of Evolutionary Biology,* 22, 2332–2341.

Mani, G. S., & Clarke, B. C. (1990). Mutational order: a major stochastic process in evolution. *Proceedings of the Royal Society of London. B. Biological Sciences*, 240(1297), 29-37.

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology and Evolution* 34(6), 531-544.

Martin, C. H. and Genner, M. J. (2009) High niche overlap between two successfully coexisting pairs of Lake Malawi cichlid fishes, *Canadian Journal of Fisheries and Aquatic Sciences*, 66(4), 579–588.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10-12.

Martin, C. H. (2013). Strong assortative mating by diet, color, size, and morphology but limited progress toward sympatric speciation in a classic example: Cameroon crater lake cichlids. *Evolution,* 67(7), 2114-2123.

Martin, C. H., Cutler, J. S., Friel, J. P., Dening Touokong, C., Coop, G., & Wainwright, P. C. (2015). Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, 69(6), 1406-1422.

Matagi, S. V. (2004). A biodiversity assessment of the Flamingo Lakes of eastern Africa. *Biodiversity*, 5(1), 13-26.

Matsumoto, A., Kasai, H., Matsuo, Y., Shizuri, Y., Ichikawa, N., Fujita, N., Ōmura, S. and Takahashi, Y. (2009) Ilumatobacter fluminis gen. nov., sp. nov., a novel actinobacterium isolated from the sediment of an estuary', *The Journal of General and Applied Microbiology*, 55(3), 201–205.

Matsumoto, A., Kasai, H., Matsuo, Y., Shizuri, Y., Ichikawa, N., Fujita, N., Ōmura, S. and Takahashi, Y. (2013). Ilumatobacter nonamiense sp. nov. and Ilumatobacter coccineum sp. nov., isolated from seashore sand. *International Journal of Systematic and Evolutionary Microbiology*, 63(9), 3404-3408.

Mavárez, J., and Linares, M. (2008). Homoploid hybrid speciation in animals. *Molecular Ecology,* 17, 4181–4185.

Mayr, E. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Mayr, E. (1947). Ecological factors in speciation. *Evolution*, 263-288.

McGee, M.D., Borstein, S.R., Meier, J.I., Marques, D.A., Mwaiko, S., Taabu, A., Kishe, M.A., O'Meara, B., Bruggmann, R., Excoffier, L. and Seehausen, O. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, 586(7827), 75-79.

McKaye, K. R., and Marsh, A. (1983). Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia,* 562(56), 245–248.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One,* 8(4), e61217.

McOrist, S., Gebhart, C. J., Boid, R., & Barns, S. M. (1995). Characterization of Lawsonia intracellularis gen. nov., sp. nov., the obligately intracellular bacterium of porcine proliferative enteropathy. *International Journal of Systematic and Evolutionary Microbiology,* 45(4), 820-825.

McPeek, M. A., & Wellborn, G. A. (1998). Genetic variation and reproductive isolation among phenotypically divergent amphipod populations. *Limnology and Oceanography*, 43(6), 1162-1169.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017a). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, 8(1), 14363.

Meier, J.I., Sousa, V.C., Marques, D.A., Selz, O.M., Wagner, C.E., Excoffier, L. and Seehausen, O. (2017b). Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Molecular Ecology,* 26, 123–141.

Meier, J.I., Salazar, P.A., Kučka, M., Davies, R.W., Dréau, A., Aldás, I., Box Power, O., Nadeau, N.J., Bridle, J.R., Rolian, C. and Barton, N.H. (2021). Haplotype tagging reveals parallel

formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences*, 118(25), e2015005118.

Meisner, J., and Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, 210, 719–731.

Mejía, O., Sánchez-Quinto, A., Gómez-Acata, E.S., Pérez-Miranda, F. and Falcón, L.I. (2022). Unraveling the Gut Microbiome of the Genus Herichthys (Pisces: Cichlidae): What Can We Learn from Museum Specimens?. *Current Microbiology*, 79(11), 346-358.

Melack, J.M. (1996). Saline and freshwater lakes of the Kenyan Rift Valley. *East African Ecosystems and their Conservation*, 7, 171-182.

Mendelson, T. C., & Safran, R. J. (2021). Speciation by sexual selection: 20 years of progress. *Trends in Ecology and Evolution,* 36(12), 1153-1163.

Mengistou, S. (2016). Invertebrates of East African Soda Lakes. *Soda Lakes of East Africa*, 205–226. Springer.

Mérot, C., Frérot, B., Leppik, E., & Joron, M. (2015). Beyond magic traits: Multimodal mating cues in Heliconius butterflies. *Evolution*, 69(11), 2891-2904.

Merrill, R.M., Dasmahapatra, K.K., Davey, J.W., Dell'Aglio, D.D., Hanly, J.J., Huber, B., Jiggins, C.D., Joron, M., Kozak, K.M., Llaurens, V. and Martin, S.H. (2015). The diversification of Heliconius butterflies: what have we learned in 150 years? *Journal of Evolutionary Biology*, 28, 1417–1438.

Mikhodyuk, O. S., Zavarzin, G. A. and Ivanovsky, R. N. (2008). Transport systems for carbonate in the extremely natronophilic cyanobacterium Euhalothece sp. *Microbiology*, 77(4), 412–418.

Miller, M. L. and Koburger, J. A. (1986). Tolerance of Plesiomonas shigelloides to pH, Sodium Chloride and Temperature. *Journal of Food Protection*, 49(11), 877–879.

Miller, A. K., Westlake, C. S., Cross, K. L., Leigh, B. A., and Bordenstein, S. R. (2021). The microbiome impacts host hybridization and speciation. *PLoS Biology,* 19, e3001417.

Mirete, S., Morgante, V., & González-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*, 38, 143-149.

Miyake, S., Ngugi, D. K., and Stingl, U. (2015). Diet strongly influences the gut microbiota of surgeonfishes. *Molecular Ecology,* 24, 656–672.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on Earth and in the ocean?. *PLoS Biology*, 9(8), e1001127.

Moreira-Silva, J., Tsui, T. K. N., Coimbra, J., Vijayan, M. M., Ip, Y. K., & Wilson, J. M. (2010). Branchial ammonia excretion in the Asian weatherloach Misgurnus anguillicaudatus. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 151(1), 40-50.

Morishita, H., & Okamoto, A. (1981). The study of salt resistance in aquatic microorganisms: Free amino acid accumulation in Aquaspirillum itersonii KF 9 grown in a saline environment. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, 21(3), 1415-1420.

Morrison, D. A. (2014). Phylogenetic networks: a review of methods to display evolutionary history. *Annual Research & Review in Biology*, 1518-1543.

Morse, S. A. (2011). Neisseria, Moraxella, Kingella and Eikenella. In *Medical Microbiology*. University of Texas Medical Branch at Galveston

Moser, F. N., Rijssel, J. C. Van, Mwaiko, S., Meier, J. I., Ngatunga, B., and Seehausen, O. (2018). The onset of ecological diversification 50 years after colonization of a crater lake by haplochromine cichlid fishes. *Proceedings of the Royal Society B*, 285(1884)

Murali, A., Bhargava, A., & Wright, E. S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, 6(1), 1-14.

Muschick, M., Indermaur, A., & Salzburger, W. (2012). Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, 22(24), 2362-2368.

Muschick, M., Nosil, P., Roesti, M., Dittmann, M.T., Harmon, L. and Salzburger, W. (2014). Testing the stages model in the adaptive radiation of cichlid fishes in East African Lake Tanganyika. *Proceedings of the Royal Society B: Biological Sciences*, 281(1795), 20140605.

Mwaniki, P., Taita, T., Sierens, T. and Triest, L. (2019). Barriers to genetic connectivity of smooth flatsedge (Cyperus laevigatus) among alkaline-saline lakes of Eastern Rift Valley (Kenya). *Aquatic Botany*, 155, 38-44.

Nagl, S., Tichy, H., Mayer, W.E., Samonte, I.E., McAndrew, B.J. and Klein, J. (2001). Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 20(3), 361-374.

Nalley, E. M., Donahue, M. J., & Toonen, R. J. (2022). Metabarcoding as a tool to examine cryptic algae in the diets of two common grazing surgeonfishes, Acanthurus triostegus and A. nigrofuscus. *Environmental DNA*, 4(1), 135-146.

Narahara, A., Bergman, H. L., Laurent, P., Maina, J. N., Walsh, P. J., & Wood, C. M. (1996). Respiratory physiology of the Lake Magadi tilapia (Oreochromis alcalicus grahami), a fish adapted to a hot, alkaline, and frequently hypoxic environment. *Physiological Zoology*, 69(5), 1114-1136.

Nevado, B., Fazalova, V., Backeljau, T., Hanssens, M., & Verheyen, E. (2011). Repeated unidirectional introgression of nuclear and mitochondrial DNA between four congeneric Tanganyikan cichlids. *Molecular Biology and Evolution*, 28(8), 2253-2267.

Newman, M. E. (1997). A model of mass extinction. *Journal of Theoretical Biology*, 189(3), 235-252.

Nonga, H.E., Mdegela, R.H., Sandvik, M., Lie, E., Miles, C.O. and Skaare, J.U. (2011) Cyanobacteria and cyanobacterial toxins in the alkaline-saline Lakes Natron and Momela, Tanzania, *Tanzania Veterinary Association Proceedings*, 34, 108–116.

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, 17, 2103–2106.

Nosil, P. (2012). *Ecological Speciation*. Oxford University Press

Nosil, P., and Feder, J. L Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 332-342.

Nyalungu, N. P., & Couldridge, V. (2020). Female mate choice and species recognition between two closely related cichlid fish of Lake Malawi, Metriaclima estherae and M. callainos. *Journal of Fish Biology*, 97(1), 75-82.

Oduor, S. O. and Schagerl, M. (2007) Phytoplankton primary productivity characteristics in response to photosynthetically active radiation in three Kenyan Rift Valley saline –alkaline lakes, *Journal of Plankton Research*, 29(12) 1041–1050.

Ofek, T., Lalzar, M., Laviad-Shitrit, S., Izhaki, I., & Halpern, M. (2021). Comparative study of intestinal microbiota composition of six edible fish species. *Frontiers in Microbiology*, 12, 760266.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. and Oksanen, M.J. (2022). *vegan*: Community Ecology Package.

Olave, M., and Meyer, A. (2020). Implementing Large Genomic Single Nucleotide Polymorphism Data Sets in Phylogenetic Network Reconstructions: A Case Study of Particularly Rapid Radiations of Cichlid Fish. *Systematic Biology*, 69(5). 848–862.

Ottenburghs, J. (2020). Ghost Introgression: Spooky Gene Flow in the Distant Past. *BioEssays,* 42, 2000012.

Owen, R.B., Muiruri, V.M., Lowenstein, T.K., Renaut, R.W., Rabideaux, N., Luo, S., Deino, A.L., Sier, M.J., Dupont-Nivet, G., McNulty, E.P. and Leet, K.,(2018). Progressive aridification in East Africa over the last half million years and implications for human evolution. *Proceedings of the National Academy of Sciences*, 115(44), 11174-11179.

Owen, R.B., Renaut, R.W., Muiruri, V.M., Rabideaux, N.M., Lowenstein, T.K., McNulty, E.P., Leet, K., Deocampo, D., Luo, S., Deino, A.L. and Cohen, A. (2019). Quaternary history of the Lake Magadi Basin, southern Kenya Rift: Tectonic and climatic controls. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 518, 97-118.

Panhuis, T. M., Butlin, R., Zuk, M., and Tregenza, T. (2001). Sexual selection and speciation. *Trends in Ecology and Evolution* 16, 364–371.

Parker, G. A., & Partridge, L. (1998). Sexual conflict and speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1366), 261-274.

Parte, A. C., Carbasse, J. S., Meier-Kolthoff, J. P., Reimer, L. C., & Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, 70(11), 5607.

Patel, A. K., Singhania, R. R., Pandey, A., Joshi, V. K., Nigam, P. S., & Soccol, C. R. (2014). Enterobacteriaceae, Coliforms and E. coli. In *Encyclopedia of Food Microbiology: Second Edition* (pp. 659-666). Elsevier Inc.

Pillonel, T., Bertelli, C., Aeby, S., de Barsy, M., Jacquier, N., Kebbi-Beghdadi, C., Mueller, L., Vouga, M. and Greub, G., (2019). Sequencing the obligate intracellular Rhabdochlamydia helvetica within its tick host Ixodes ricinus to investigate their symbiotic relationship. *Genome Biology and Evolution*, 11(4), 1334-1344.

Pedersen, T. L. (2020a). *tidygraph*: *a tidy API for graph manipulation*. R package version, 1.2.1.

Pedersen, T. (2022b). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. R package version, 2.0.5.

Penk, S.B., Altner, M., F. Cerwenka, A., Schliewen, U.K. and Reichenbacher, B., (2019). New fossil cichlid from the middle Miocene of East Africa revealed as oldest known member of the Oreochromini. *Scientific Reports*, 9(1), 10198.

Péter, A., 2011. Solomon Coder: a simple solution for behavior coding. Version 11.01. 22.

Pfennig, K., & Pfennig, D. (2009). Character displacement: ecological and reproductive responses to a common evolutionary problem. *The Quarterly Review of Biology*, 84(3), 253-276.

Philippart, J.C., Ruwet, J.C. (1982). *The Biology and Culture of Tilapia*. ICLARM-International Center for Living Aquatic Resources Management, Manila, Philippines.

Plenderleith, M., Oosterhout, C.V., Robinson, R.L. and Turner, G.F., (2005). Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. *Biology Letters*, 1(4), 411-414.

Poelstra, J. W., Richards, E. J., & Martin, C. H. (2018). Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Molecular Ecology*, 27(21), 4270-4288.

Pörtner, H.O., Schulte, P.M., Wood, C.M. and Schiemer, F. (2010). Niche dimensions in fishes: an integrative view. *Physiological and Biochemical Zoology*, 83(5), 808-826.

Prabu, E., Rajagopalsamy, C. B. T., Ahilan, B., Jeevagan, I. J. M. A., & Renuhadevi, M. (2019). Tilapia–an excellent candidate species for world aquaculture: a review. *Annual Research & Review in Biology*, 1-14.

Presgraves, D.C., Balagopalan, L., Abmayr, S.M. and Orr, H.A.,(2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature*, 423(6941), 715-719.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS One*, 5(3), e9490.

Pujalte, M.J., Lucena, T., Ruvira, M.A., Arahal, D.R. and Macián, M.A. (2014). The Family Rhodobacteraceae. In *The Prokaryotes*. Springer.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafferty, N.E. and Boughman, J.W. (2006). Olfactory mate recognition in a sympatric species pair of three-spined sticklebacks. *Behavioral Ecology*, 17(6), 965-970.

Rajkov, J., Weber, A. A. T., Salzburger, W., & Egger, B. (2018). Immigrant and extrinsic hybrid inviability contribute to reproductive isolation between lake and river cichlid ecotypes. *Evolution*, 72(11), 2553-2564.

Randall, D.J., Wood, C.M., Perry, S.F., Bergman, H.L., Maloiy, G.M.O., Mommsen, T.P. and Wright, P.A. (1989). Urea excretion as a strategy for survival in a fish living in a very alkaline environment. *Nature*, 337(6203), 165-166.

Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B. and Westram, A.M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450-1477.

Rennison, D.J., Rudman, S.M. and Schluter, D (2019). Parallel changes in gut microbiome composition and function during colonization, local adaptation and ecological speciation. *Proceedings of the Royal Society B*, 286(1916), 20191911.

Ribbink, A. J., and Hill, B. J. (1979). Depth equilibration by two predatory cichlid fish from Lake Malawi. *Journal of Fish Biology* 14, 507–510.

Rice, W. R. (1998). “Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproduction isolation,” in *Endless forms: Species and Speciation*, 261–270. Oxford University Press.

Richards, E., Servedio, M., and Martin, C. (2019). Searching for sympatric speciation in the genomic era. *BioRxiv*, 367623.

Rico, C., and Turner, G. F. (2002). Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Molecular Ecology*, 11, 1585–1590.

Riera, J. L., & Baldo, L. (2020). Microbial co-occurrence networks of gut microbiota reveal community conservation and diet-associated shifts in cichlid fishes. *Animal Microbiome*, 2, 1-13.

Riesch, R., Tobler, M., & Plath, M. (2015). *Extremophile fishes: Ecology, Evolution, and Physiology of Teleosts in Extreme Environments*. Springer.

Ringø, E.Z.Z.V., Zhou, Z., Vecino, J.G., Wadsworth, S., Romero, J., Krogdahl, Å., Olsen, R.E., Dimitroglou, A., Foey, A., Davies, S. and Owen, M. (2016). Effect of dietary components on the gut microbiota of aquatic animals. A never-ending story?. *Aquaculture Nutrition*, 22(2), 219-282.

Ritchie, M.G., 2007. Sexual selection and speciation. *Annual Review of Ecology, Evolution, and Systematic*., 38, 79-102.

Recknagel, H., Elmer, K. R., & Meyer, A. (2014). Crater lake habitat predicts morphological diversity in adaptive radiations of cichlid fishes. *Evolution*, 68(7), 2145-2155.

Rieseberg, L. H., Archer, M. A., and Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity,* 83, 363–372.

Roberts, N., Taieb, M., Barker, P., Damnati, B., Icole, M., & Williamson, D. (1993). Timing of the Younger Dryas event in East Africa from lake-level changes. *Nature*. 366(6451), 146-148.

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology,* 28, 4737–4754.

Roh, S.W., Nam, Y.D., Chang, H.W., Sung, Y., Kim, K.H., Lee, H.J., Oh, H.M. and Bae, J.W. (2007). Natronococcus jeotgali sp. nov., a halophilic archaeon isolated from shrimp jeotgal, a traditional fermented seafood from Korea. *International Journal of Systematic and Evolutionary Microbiology*, 57(9), 2129-2131.

Rometsch, S. J., Torres-Dowdall, J., & Meyer, A. (2020). Evolutionary dynamics of pre-and postzygotic reproductive isolation in cichlid fishes. *Philosophical Transactions of the Royal Society B*, 375(1806), 20190535.

Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H.H., El Taher, A., Indermaur, A., Malinsky, M., Ricci, V., Kahmen, A. and Jentoft, S. (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, 589(7840), 76-81.

Rosser, N., Queste, L.M., Cama, B., Edelman, N.B., Mann, F., Mori Pezo, R., Morris, J., Segami, C., Velado, P., Schulz, S. and Mallet, J.L. (2019). Geographic contrasts between pre-and postzygotic barriers are consistent with reinforcement in Heliconius butterflies. *Evolution*, 73(9), 1821-1838.

Rosshart, S.P., Vassallo, B.G., Angeletti, D., Hutchinson, D.S., Morgan, A.P., Takeda, K., Hickman, H.D., McCulloch, J.A., Badger, J.H., Ajami, N.J. and Trinchieri, G. (2017). Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell*, 171(5), 1015-1028.

Rundle, H. D., & Schluter, D. (2004). Natural selection and ecological speciation in sticklebacks. *Adaptive Speciation*, 19(3), 192-209.

Rundle, H. D., and Nosil, P. (2005). Ecological speciation. *Ecology Letters,* 8(3), 336–352.

Ryan, M.J. and Cummings, M.E. (2013). Perceptual biases and mate choice. *Annual Review of Ecology, Evolution, and Systematics*, 44, 437-459.

Safran, R.J., Scordato, E.S., Symes, L.B., Rodríguez, R.L. and Mendelson, T.C. (2013). Contributions of natural and sexual selection to the evolution of premating reproductive isolation: a research agenda. *Trends in Ecology & Evolution*, 28(11).

Salzburger, W., Baric, S., & Sturmbauer, C. (2002). Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, 11(3), 619-625.

Salzburger, W., & Meyer, A. (2004). The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften*, 91, 277-290.

Salzburger, W. (2009). The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Spixiana* 37, 135–149.

Salzburger, W., Van Bocxlaer, B., & Cohen, A. S. (2014). Ecology and evolution of the African Great Lakes and their faunas. *Annual Review of Ecology, Evolution, and Systematics*, 45, 519-545.

Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews Genetics.* 19, 705–717.

Santos, J. A., Rodríguez-Calleja, J. M., Otero, A., & García-López, M. L. (2015). "Plesiomonas". In *Molecular Medical Microbiology* (pp. 1111-1123). Academic Press.

Santos, M. E., Lopes, J. F., and Kratochwil, C. F. (2023). East African cichlid fishes. *Evodevo,* 14(1).

Savolainen, V., Anstett, M.C., Lexer, C., Hutton, I., Clarkson, J.J., Norup, M.V., Powell, M.P., Springate, D., Salamin, N. and Baker, W.J. (2006). Sympatric speciation in palms on an oceanic island. *Nature*, 441, 210–213.

Schagerl, M., Burian, A., Gruber-Dorninger, M., Oduor, S. O., & Kaggwa, M. N. (2015). Phytoplankton succession of African soda lakes with a special focus on Arthrospira fusiformis. *Fottea*, 15, 245-257.

Schagerl, M. (2016). *Soda lakes of East Africa*. Springer.

Sherwood, A. R., & Presting, G. G. (2007). Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria 1. *Journal of Phycology*, 43(3), 605-608.

Schliewen, U. K., & Klee, B. (2004). Reticulate sympatric speciation in Cameroonian crater lake cichlids. *Frontiers in Zoology*, 1(1), 1-12.

Schliewen, U., Rassmann, K., Markmann, M., Markert, J., Kocher, T., and Tautz, D. (2001). Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology*, 10, 1471–1488.

Schnürer, A., Schink, B. and Svensson, B. H. (1996). Clostridium ultunense sp. nov., a mesophilic bacterium oxidizing acetate in syntrophic association with a hydrogenotrophic methanogenic bacterium, *International Journal of Systematic Bacteriology*, 46(4), 1145–1152.

Schluter, D. (1993). Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. *Ecology*, 74, 699–709.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution* 16, 372–380.

Schoener, T. W. (1974). Resource Partitioning in Ecological Communities, *Science*, 185(4145), 27–39.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18(2), 233-234.

Scordato, E. S., Symes, L. B., Mendelson, T. C., & Safran, R. J. (2014). The role of ecology in speciation by sexual selection: a systematic empirical review. *Journal of Heredity*, 105(S1), 782-794.

Seck, E. H., Dufour, J. C., Raoult, D., & Lagier, J. C. (2018). Halophilic & halotolerant prokaryotes in humans. *Future Microbiology*, 13(07), 799-812.

Seegers L., Tichy ,H. (1999) The Oreochromis alcalicus flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. Ichthyological Explorations of Freshwaters, 10, 97–146.

Seegers, L., Sonnenberg, R., & Tichy, H. (2001). The Alcolapia group, a remarkable species flock from lakes Natron, Tanzania, and Magadi, Kenya: a further piece of the puzzle of cichlid evolution. *Journal of Aquariculture and Aquatic Sciences*, 9, 335-364.

Seegers, L. (2008). The fishes collected by GA Fischer in East Africa in 1883 and 1885/86. *Zoosystematics and Evolution*, 84(2), 149-195.

Seehausen, O. (1996). *Lake Victoria rock cichlids: Taxonomy, Ecology and Distribution*. Verduyn cichlids.

Seehausen, O., van Alphen, J. J. M., & Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science,* 277(5333), 1808-1811.

Seehausen, O., & van Alphen, J. J. (1998). The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behavioral Ecology and Sociobiology*, 42, 1-8.

Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597), 1987-1998.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D., Miyagi, R., Van Der Sluijs, I., Schneider, M.V., Maan, M.E., Tachida, H. and Imai, H. (2008). Speciation through sensory drive in cichlid fish. *Nature*, 455(7213), 620-626.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.P., Bank, C., Brännström, Å. and Brelsford, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176-192.

Seehausen, O., & Wagner, C. E. (2014). Speciation in freshwater fishes. *Annual Review of Ecology, Evolution, and Systematics*, 45, 621-651.

Seehausen, O. (2015). Process and pattern in cichlid radiations – inferences for understanding unusually high rates of evolutionary diversification. *New Phytologist*, 207, 304–312.

Sefc, K.M., Hermann, C.M., Steinwender, B., Brindl, H., Zimmermann, H., Mattersdorfer, K., Postl, L., Makasa, L., Sturmbauer, C. and Koblmüller, S., (2015). Asymmetric dominance and asymmetric mate choice oppose premating isolation after allopatric divergence. *Ecology and Evolution* 5, 1549–1562.

Sefc, K. M., Mattersdorfer, K., Hermann, C. M., and Koblmüller, S. (2017). Past lake shore dynamics explain present pattern of unidirectional introgression across a habitat barrier. *Hydrobiologia* 791, 69–82.

Selz, Oliver M., Pierotti, M. E. R., Maan, M. E., Schmid, C., & Seehausen, O. (2014). Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioural Ecology*, 25(3), 612-626.

Selz, O. M., Thommen, R., Pierotti, M. E. R., Anaya-Rojas, J. M., & Seehausen, O. (2016). Differences in male coloration are predicted by divergent sexual selection between populations of a cichlid fish. *Proceedings of the Royal Society B: Biological Sciences*, 283(1830), 20160172.

Sepulveda, J. and Moeller, A. H. (2020) The Effects of Temperature on Animal Gut Microbiomes, *Frontiers in Microbiology*, 11, 384.

Servedio, M. R., and Noor, M. A. F. (2003). The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, 339–364.

Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: 'magic' but not rare? *Trends in Ecology and Evolution* 26(8), 389-397.

Servedio, M. R. (2016). Geography, assortative mating, and the effects of sexual selection on speciation with gene flow. *Evolutionary Applications*, 9(1), 91-102.

Servedio, M.R. and Boughman, J.W., (2017). The role of sexual selection in local adaptation and speciation. *Annual Review of Ecology, Evolution, and Systematics*, 48, 85-109.

Servedio, M. R., and Bürger, R. (2020). The effectiveness of pseudomagic traits in promoting divergence and enhancing local adaptation. *Evolution*, 74, 2438–2450.

Shah, H. N. (1992) The Genus Porphyromonas, In *The Prokaryotes*, 3608–3619. Springer.

Shoo, R. A. (2020). Ecotourism Potential and Challenges at Lake Natron Ramsar Site, Tanzania. Protected Areas in Northern Tanzania: Local Communities, *Land Use Change, and Management Challenges*, 75-90.

Shropshire, J. D., and Bordenstein, S. R. (2016). Speciation by symbiosis: The microbiome and behavior. *MBio* 7(2).

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787-792.

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, 20(24), 5123-5140.

Smith, C.C., Snowberg, L.K., Gregory Caporaso, J., Knight, R. and Bolnick, D.I. (2015). Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota, *The ISME Journal*, 9(11), 2515–2526.

Srinivas, T.N.R., Anil Kumar, P., Sasikala, C., Ramana, C.V. and Imhoff, J.F. (2007) Rhodobacter vinaykumarii sp. nov., a marine phototrophic alphaproteobacterium from tidal waters, and emended description of the genus Rhodobacter, *International Journal of Systematic and Evolutionary Microbiology*, 57(9), 1984–1987.

Stager, J. C., and Johnson, T. C. (2008). The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia,* 596, 5–16.

Stankowski, S., & Streisfeld, M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences*, 282(1814), 20151666.

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*. 75(6), 1256-1273.

Stauffer Jr, J. R. and McKaye, K. R. (2002). Descriptions of three new species of cichlid fishes (Teleostei: Cichlidae) from Lake Xiloá, Nicaragua, *Cuadernos de Investigación*, (12), 1–18.

Stelkens, R. B., & Seehausen, O. (2009). Phenotypic divergence but not genetic distance predicts assortative mating among species of a cichlid fish radiation. *Journal of Evolutionary Biology,* 22(8), 1679-1694.

Stelkens, R. B., Young, K. A., & Seehausen, O. (2010). The accumulation of reproductive incompatibilities in African cichlid fish. *Evolution*, 64(3), 617-633.

Streelman, J. T. and Danley, P. D. (2003). The stages of vertebrate evolutionary radiation, *Trends in Ecology & Evolution*, 18(3), 126–131.

Stroud, J.T. and Losos, J.B. (2016). Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, 47, 507-532.

Sturmbauer, C., Baric, S., Salzburger, W., Rüber, L., & Verheyen, E. (2001). Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution*, 18(2), 144-154.

Svardal, H., Salzburger, W., & Malinsky, M. (2021). Genetic variation and hybridization in evolutionary radiations of cichlid fishes. *Annual Review of Animal Biosciences*, 9, 55-79.

Sweetlove, L. (2011). Number of species on Earth tagged at 8.7 million. *Nature*, 23.

Taborsky, B., Guyer, L., & Demus, P. (2014). Prudent habitat choice: a novel mechanism of size-assortative mating. *Journal of Evolutionary Biology* 27(6), 1217-1228.

Takai, K., Moser, D. P., Onstott, T. C., Spoelstra, N., Pfiffner, S. M., Dohnalkova, A., & Fredrickson, J. K. (2001). Alkaliphilus transvaalensis gen. nov., sp. nov., an extremely alkaliphilic bacterium isolated from a deep South African gold mine. *International Journal of Systematic and Evolutionary Microbiology*, 51(4), 1245-1256.

Takeuchi, M., Katayama, T., Yamagishi, T., Hanada, S., Tamaki, H., Kamagata, Y., Oshima, K., Hattori, M., Marumo, K., Nedachi, M. and Maeda, H. (2014). Methyloceanibacter caenitepidi gen. nov., sp. nov., a facultatively methylotrophic bacterium isolated from marine sediments near a hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology*, 64(2), 462-468.

Takeyama, T. (2021). Feeding ecology of lake Tanganyika cichlids. *The Behavior, Ecology and Evolution of Cichlid Fishes*, 715-751.

Tebbs, E. J., Remedios, J. J., Avery, S. T., & Harper, D. M. (2013). Remote sensing the hydrological variability of Tanzania's Lake Natron, a vital Lesser Flamingo breeding site under threat. *Ecohydrology & Hydrobiology*, 13(2), 148-158.

Templeton, A. R. (1981). Mechanisms of speciation-a population genetic approach. *Annual Review of Ecology and Systematics*, 12(1), 23-48.

Templeton, A. R. (1989). The meaning of species and speciation: a genetic perspective. In *Speciation and its Consequences*, 3-27, Sinauer.

Terai, Y., Seehausen, O., Sasaki, T., Takahashi, K., Mizoiri, S., Sugawara, T., Sato, T., Watanabe, M., Konijnendijk, N., Mrosso, H.D.J. and Tachida, H., (2006). Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biology*, 4(12), p.e433.

Terhorst, J., & Song, Y. S. (2015). Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proceedings of the National Academy of Sciences*, 112(25), 7677-7682.

Teuber, M. (1995) The genus Lactococcus, In *The Genera of Lactic Acid Bacteria*, 173–234. Springer.

Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9, 1–16.

Thomas, V., Casson, N. and Greub, G. (2006) Criblamydia sequanensis, a new intracellular Chlamydiales isolated from Seine river water using amoebal co-culture, *Environmental Microbiology*, 8(12), 2125–2135.

Tichy, H., and Seegers, L. (1999). The Oreochromis alcalicus flock (Teleostei: Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: a model for the evolution of “new” species flocks in historical times? *Ichthyological Exploration of Freshwaters* 10, 147–174.

Tindall, B. J., Ross, H. N. M. and Grant, W. D. (1984) Natronobacterium gen. nov. and Natronococcus gen. nov., Two New Genera of Haloalkaliphilic Archaebacteria, *Systematic and Applied Microbiology*, 5(1), 41–57.

Tinghitella, R. M., Lackey, A. C., Durso, C., Koop, J. A., & Boughman, J. W. (2020). The ecological stage changes benefits of mate choice and drives preference divergence. *Philosophical Transactions of the Royal Society B*, 375(1806), 20190546.

Trewavas, E. (1983). *Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia*. Natural History Museum Library, London.

Tsuchiya, C., Sakata, T. and Sugita, H. (2008) Novel ecological niche of Cetobacterium somerae, an anaerobic bacterium in the intestinal tracts of freshwater fish, *Letters in Applied Microbiology*, 46(1), 43–48.

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology & Evolution*, 16(7), 330-343.

Turner, G. F., Seehausen, O., Knight, M. E., Allender, C. J., & Robinson, R. L. (2001). How many species of cichlid fishes are there in African lakes? *Molecular Ecology*, 10(3), 793-806.

Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic Islands of Speciation in Anopheles gambiae. *PLoS Biol,* 3, e285.

Tyers, A. M., Bavin, D., Cooke, G. M., Griggs, C., & Turner, G. F. (2014). Peripheral isolate speciation of a Lake Malawi cichlid fish from shallow muddy habitats. *Evolutionary Biology*, 41, 439-451.

Vandekerckhove, T. T., Willems, A., Gillis, M., & Coomans, A. (2000). Occurrence of novel verrucomicrobial species, endosymbiotic and associated with parthenogenesis in Xiphinema americanum-group species (Nematoda, Longidoridae). *International Journal of Systematic and Evolutionary Microbiology*, 50(6), 2197-2205.

van Der Sluijs, I., Van Dooren, T. J., Hofker, K. D., van Alphen, J. J., Stelkens, R. B., & Seehausen, O. (2008). Female mating preference functions predict sexual selection against hybrids between sibling species of cichlid fish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505), 2871-2877.

Van Doorn, G. S., Edelaar, P., and Weissing, F. J. (2009). On the origin of species by natural and sexual selection. *Science,* (326)5960 326, 1704–1707.

Van Steenberge, M., Jublier, N., Kéver, L., Gresham, S., Derycke, S., Snoeks, J., Parmentier, E., Poncin, P. and Verheyen, E. (2022). The initial response of females towards congeneric males matches the propensity to hybridise in Ophthalmotilapia. *bioRxiv*. 2021-2008.

Vareschi, E. and Vareschi, A. (1979). The Ecology of Lake Nakuru (Kenya): II. Biomass and Spatial Distribution of Fish (Tilapia Grahami Boulenger=Sarotherodon Alcalicum Grahami Boulenger), *Oecologia*, 37(3), 321–335.

Vekeman, B., Kerckhof, F.M., Cremers, G., De Vos, P., Vandamme, P., Boon, N., Op den Camp, H.J. and Heylen, K. (2016). New Methyloceanibacter diversity from North Sea sediments includes methanotroph containing solely the soluble methane monooxygenase, *Environmental Microbiology*, 18(12), 4523–4536.

Vernaz, G., Hudson, A.G., Santos, M.E., Fischer, B., Carruthers, M., Shechonge, A.H., Gabagambi, N.P., Tyers, A.M., Ngatunga, B.P., Malinsky, M. and Durbin, R. (2022). Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. *Nature Ecology & Evolution,* 612(6), 1940–1951.

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, 106, 9939-9946.

Vincens, A., & Casanova, J. (1987). Modern background of Natron-Magadi basin (Tanzania-Kenya): physiography, climate, hydrology and vegetation. *Sciences Géologiques, bulletins et mémoires*, 40(1), 9-21.

Wagner, C. E., Harmon, L. J., and Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature,* 487, 366–369.

Wallbank, R.W., Baxter, S.W., Pardo-Diaz, C., Hanly, J.J., Martin, S.H., Mallet, J., Dasmahapatra, K.K., Salazar, C., Joron, M., Nadeau, N. and McMillan, W.O. (2016). Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biology*, 14(1), e1002353.

Walsh P.J., Grosell M.A., Goss G.G., Bergman H.L., Bergman A.N., Wilson P.A., Laurent P.I., Alper S.L., Smith C.P., Kamunde C.O., Wood C.M. (2001). Physiological and molecular characterization of urea transport by the gills of the Lake Magadi tilapia (Alcolapia grahami). *Journal of Experimental Biology*, 204(3), 509-520.

Wang, A. R., Ran, C., Ringø, E., & Zhou, Z. G. (2018). Progress in fish gastrointestinal microbiota research, *Reviews in Aquaculture*, 10(3), 626–640.

Weiss, J. D., Cotterill, F. P. D., and Schliewen, U. K. (2015). Lake Tanganyika—A "Melting Pot" of Ancient and Young Cichlid Lineages (Teleostei: Cichlidae)? *PLoS One* 10, e0125043.

Wellborn, G. A., and Langerhans, R. B. (2015). Ecological opportunity and the adaptive diversification of lineages. *Ecology and Evolution,* 5, 176–195.

Warton, D. I., Wright, S. T., & Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, 3(1), 89-101.

White, L. J., Sutton, G., Shechonge, A., Day, J. J., Dasmahapatra, K. K., & Pownall, M. E. (2020). Adaptation of the carbamoyl-phosphate synthetase enzyme in an extremophile fish. *Royal Society Open Scienc*e, 7(10), 201200.

White, L.J., Rose, M., Lawson, M., Joyce, D., Smith, A.M., Thomas, G.H., Dasmahapatra, K.K. and Pownall, M.E. (2022). Two closely related ureotelic fish species of the genus Alcolapia express different levels of ammonium transporters in gills. *Biology Open,* 11. bio059575.

Wickham, H. (2016). Data analysis. *ggplot2*. Springer.

Willems, A. and Collins, M. D. (1994) Phylogenetic placement of Sarcina ventriculi and Sarcina maxima within group I clostridium, a possible problem for future revision of the genus Clostridium, *International Journal of Systematic Bacteriology*, 44(3), 591–593.

Williamson, D., Taieb, M., Damnati, B., Icole, M., & Thouveny, N. (1993). Equatorial extension of the Younger Dryas event: rock magnetic evidence from Lake Magadi (Kenya). *Global and Planetary Change*, 7(1-3), 235-242.

Wilson, P. J., Wood, C. M., Maina, J. N., and White, B. N. (2000). Genetic structure of Lake Magadi tilapia populations. *Journal of Fish Biology* 56, 590–603.

Wilson, P. J., Wood, C. M., Walsh, P. J., Bergman, A. N., Bergman, H. L., Laurent, P., & White, B. N. (2004). Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, 77(4), 537-555.

Winkelmann, K., Genner, M. J., Takahashi, T., & Rüber, L. (2014). Competition-driven speciation in cichlid fish. *Nature Communications*, 5(1), 3.

Wirtz, P. (1999). Mother species–father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, 58(1), 1-12.

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87-100.

Wolterink, A., Kim, S., Muusse, M., Kim, I.S., Roholl, P.J., van Ginkel, C.G., Stams, A.J. and Kengen, S.W. (2005). Dechloromonas hortensis sp. nov. and strain ASK-1, two novel (per)chlorate-reducing bacteria, and taxonomic description of strain GR-1, *International Journal of Systematic and Evolutionary Microbiology*, 55(5), 2063–2068.

Wong, S., & Rawls, J. F. (2012). Intestinal microbiota composition in fishes is influenced by host ecology and environment. *Molecular Ecology*. (919) 966-6927.

Wong, B. T. and Lee, D. J. (2014) Pseudomonas yangmingensis sp. nov., an alkaliphilic denitrifying species isolated from a hot spring, *Journal of Bioscience and Bioengineering*, 117(1), 71–74.

Woo, P.C., Teng, J.L., Leung, K.W., Lau, S.K., Woo, G.K., Wong, A.C., Wong, M.K. and Yuen, K.Y. (2005). Anaerospora hongkongensis Gen. Nov. Sp. Nov., a Novel Genus and Species with Ribosomal DNA Operon Heterogeneity Isolated from an Intravenous Drug Abuser with Pseudobacteremia, *Microbiology and Immunology*, 49(1), 31–39.

Wood, C. M., Bergman, H. L., Laurent, P., Maina, J. N., Narahara, A., and Walsh, P. J. (1994). Urea production, acid-base regulation and their interactions in the Lake Magadi tilapia, a unique teleost adapted to a highly alkaline environment. *Journal of Experimental Biology,* 189, 13–36.

Wood, C. M., Wilson, P., Bergman, H. L., Bergman, A. N., Laurent, P., Otiang'a-Owiti, G., & Walsh, P. J. (2002). Obligatory urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity and alkalinity. *Physiological and Biochemical Zoology*, 75(2), 111-122.

Wood, C.M., Bergman, H.L., Bianchini, A., Laurent, P., Maina, J., Johannsson, O.E., Bianchini, L.F., Chevalier, C., Kavembe, G.D., Papah, M.B. and Ojoo, R.O. (2012). Transepithelial potential in the Magadi tilapia, a fish living in extreme alkalinity. *Journal of Comparative Physiology B*, 182, 247-258.

Wood, C.M., Brix, K.V., De Boeck, G., Bergman, H.L., Bianchini, A., Bianchini, L.F., Maina, J.N., Johannsson, O.E., Kavembe, G.D., Papah, M.B. and Letura, K.M. (2016). Mammalian metabolic rates in the hottest fish on earth. *Scientific Reports*, 6(1), 1-9.

Woodhouse, C. W. (1912). The fish in Lake Magadi. *East African Geographical Review*, 2(4), 95-97.

Woolley, K. J. and Woolley, K. (1987). The soluble c-type cytochromes from the bacterium Aquaspirillum itersonii, *European Journal of Biochemistry*, 166(1), 131–137.

Wright, P. A., Perry, S. F., Randall, D. J., Wood, C. M., & Bergman, H. (1990). The effects of reducing water pH and total CO2 on a teleost fish adapted to an extremely alkaline environment. *Journal of Experimental Biology*, 151(1), 361-369.

Wright, E. S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. *The R Journal*, 8(1), 352.

Wright, D. S., Rietveld, E., & Maan, M. E. (2018). Developmental effects of environmental light on male nuptial coloration in Lake Victoria cichlid fish. *PeerJ*. 6, e4209.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology,* 14, 851–865.

Wu, X. Y., Shi, K. L., Xu, X. W., Wu, M., Oren, A., & Zhu, X. F. (2010). Alkaliphilus halophilus sp. nov., a strictly anaerobic and halophilic bacterium isolated from a saline lake, and emended description of the genus Alkaliphilus. *International Journal of Systematic and Evolutionary Microbiology*, 60(12), 2898-2902.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1), 1-11.

Yoon, J. H., Kang, S. J., Park, S., Lee, S. Y., & Oh, T. K. (2007). Reclassification of Aquaspirillum itersonii and Aquaspirillum peregrinum as Novispirillum itersonii gen. nov., comb. nov. and Insolitispirillum peregrinum gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 57(12), 2830-2835.

Yukilevich, R. (2012). Asymmetrical patterns of speciation uniquely support reinforcement in Drosophila. *Evolution*. 66(5), 1430-1446.

Zaccara, S., Crosa, G., Vanetti, I., Binelli, G., Harper, D.M., Mavuti, K.M., Balarin, J.D. and Britton, J.R. (2014). Genetic and morphological analyses indicate high population mixing in the endangered cichlid Alcolapia flock of East Africa. *Conservation Genetics,* 15, 429–440.

Zhang, Y.G., Chen, J.Y., Wang, H.F., Xiao, M., Yang, L.L., Guo, J.W., Zhou, E.M., Zhang, Y.M. and Li, W.J. (2016). Egicoccus halophilus gen. Nov., sp. nov., A halophilic, Alkalitolerant actinobacterium and proposal of Egicoccaceae fam. nov. and Egicoccales ord. nov., *International Journal of Systematic and Evolutionary Microbiology*, 66(2), 530–535.

Zhang, Y.G., Wang, H.F., Yang, L.L., Zhou, X.K., Zhi, X.Y., Duan, Y.Q., Xiao, M., Zhang, Y.M. and Li, W.J. (2016). Egibacter rhizosphaerae gen. Nov., sp. nov., an obligately halophilic, facultatively alkaliphilic actinobacterium and proposal of Egibaceraceae fam. nov. and Egibacterales ord. nov., *International Journal of Systematic and Evolutionary Microbiology*, 66(1), 271–275.

Zhang, C., Dong, B., Wang, R., Su, Y., Han, S., Yu, X., Zhao, Z., Fu, G., Sun, C. and Wu, M. (2017). Luteolibacter flavescens sp. nov., isolated from deep seawater. *International Journal of Systematic and Evolutionary Microbiology*, 67(3), 729-735.

Zakharyuk, A., Kozyreva, L., Ariskina, E., Troshina, O., Kopitsyn, D., & Shcherbakova, V. (2017). Alkaliphilus namsaraevii sp. nov., an alkaliphilic iron-and sulfur-reducing bacterium isolated from a steppe soda lake. *International Journal of Systematic and Evolutionary Microbiology*, 67(6), 1990-1995.

Zavarzina, D. G., Tourova, T. P., Kolganova, T. V., Boulygina, E. S., & Zhilina, T. N. (2009). Description of Anaerobacillus alkalilacustre gen. nov., sp. nov.—Strictly anaerobic diazotrophic bacillus isolated from soda lake and transfer of Bacillus arseniciselenatis, Bacillus macyae, and Bacillus alkalidiazotrophicus to Anaerobacillus as the new combinations A. arseniciselenatis comb. nov., A. macyae comb. nov., and A. alkalidiazotrophicus comb. nov. *Microbiology*, 78, 723-731.

Zhilina, T. N., Kevbrin, V. V., Tourova, T. P., Lysenko, A. M., Kostrikina, N. A., & Zavarzin, G. A. (2005). Clostridium alkalicellum sp. nov., an obligately alkaliphilic cellulolytic bacterium from a soda lake in the Baikal region. *Microbiology*, 74, 557-566.

Zhou, H., He, K., Chen, J., & Zhang, X. (2022). LinDA: linear models for differential abundance analysis of microbiome compositional data. *Genome Biology*, 23(1), 1-23.

Zhu, J., Wen, D., Yu, Y., Meudt, H. M., & Nakhleh, L. (2018). Bayesian inference of phylogenetic networks from bi-allelic genetic markers. *PLoS Computational Biology*, 14(1), e1005932.

Zouros, E. (1973). Genic differentiation associated with the early stages of speciation in the mulleri subgroup of Drosophila. *Evolution*, 27, 601-621.