Article

The ecological and genomic basis of explosive adaptive radiation

https://doi.org/10.1038/s41586-020-2652-7

Received: 14 February 2020

Accepted: 22 May 2020

Published online: 26 August 2020

Check for updates

Matthew D. McGee[1,2,3,12]✉, Samuel R. Borstein[4,12], Joana I. Meier[1,2,5], David A. Marques[1,2], Salome Mwaiko[2], Anthony Taabu[6], Mary A. Kishe[7], Brian O'Meara[8], Rémy Bruggmann[9,10], Laurent Excoffier[10,11] & Ole Seehausen[1,2]✉

Speciation rates vary considerably among lineages, and our understanding of what drives the rapid succession of speciation events within young adaptive radiations remains incomplete[1–11]. The cichlid fish family provides a notable example of such variation, with many slowly speciating lineages as well as several exceptionally large and rapid radiations[12]. Here, by reconstructing a large phylogeny of all currently described cichlid species, we show that explosive speciation is solely concentrated in species flocks of several large young lakes. Increases in the speciation rate are associated with the absence of top predators; however, this does not sufficiently explain explosive speciation. Across lake radiations, we observe a positive relationship between the speciation rate and enrichment of large insertion or deletion polymorphisms. Assembly of 100 cichlid genomes within the most rapidly speciating cichlid radiation, which is found in Lake Victoria, reveals exceptional 'genomic potential'—hundreds of ancient haplotypes bear insertion or deletion polymorphisms, many of which are associated with specific ecologies and shared with ecologically similar species from other older radiations elsewhere in Africa. Network analysis reveals fundamentally non-treelike evolution through recombining old haplotypes, and the origins of ecological guilds are concentrated early in the radiation. Our results suggest that the combination of ecological opportunity, sexual selection and exceptional genomic potential is the key to understanding explosive adaptive radiation.

Many of the geographical, ecological and genetic factors that contribute to speciation are now well-established, owing to the maturation of theoretical and empirical speciation research over the past three decades[1–8]. The origin of new species often starts in geographical isolation through the gradual accumulation of reproductive incompatibilities and/or divergent adaptation, and ecological differentiation is often completed by the displacement of ecological characters after return to sympatry[5]. Under some conditions, individual speciation events can be sudden and rapid[6,9], but current models fail to explain many rapid speciation events that occur in short succession within one lineage and region. This phenomenon occurs within large adaptive radiations, as many new species emerge in very short succession and come together in rich assemblages of ecologically diverse species that coexist sympatrically within extremely short time spans, too short for incompatibilities to emerge and far too short for new heritable variation to accumulate between successive speciation events. This phenomenon requires explanations for which neither the classical allopatric models with subsequent character displacement nor ecological speciation models are sufficient[2,6,9–11].

We study the factors that govern the dynamics of explosive diversification and accumulation of species diversity in the fish family Cichlidae, which contains more than 1,700 described, valid and extant species, comprising more than 5% of all extant teleost fishes. Approximately half of these species are endemic to lakes and are often part of young and ecologically diverse species flocks that occupy many distinct adaptive zones[12]. Other cichlid species also display a wide range of ecological and morphological diversity, although these species have diversified on much longer timescales.

Factors that influence speciation rate across all cichlids

We estimate variation in speciation rates among cichlid fishes from a fully sampled phylogeny that includes all currently taxonomically valid cichlid species described before 2019 (n = 1,712 species; Fig. 1), by combining nuclear genes, mitochondrial genes and meristic characters, and implementing topological constraints from previous phylogenomic studies of smaller clades. We estimate speciation rates using the diversification rate (DR) statistic[13,14]. Our results verify that rates

[1]Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [2]Department of Fish Ecology and Evolution, Centre for Ecology, Evolution, and Biogeochemistry, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Kastanienbaum, Switzerland. [3]School of Biological Sciences, Monash University, Melbourne, Victoria, Australia. [4]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [5]Department of Zoology, University of Cambridge, Cambridge, UK. [6]National Fisheries Resources Research Institute, Jinja, Uganda. [7]Tanzania Fisheries Research Institute, Dar es Salaam, Tanzania. [8]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA. [9]Interfaculty Bioinformatics Unit, University of Bern, Bern, Switzerland. [10]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [11]Computational and Molecular Population Genetics (CMPG), Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [12]These authors contributed equally: Matthew D. McGee, Samuel R. Borstein.
✉e-mail: matt.mcgee@monash.edu; ole.seehausen@eawag.ch

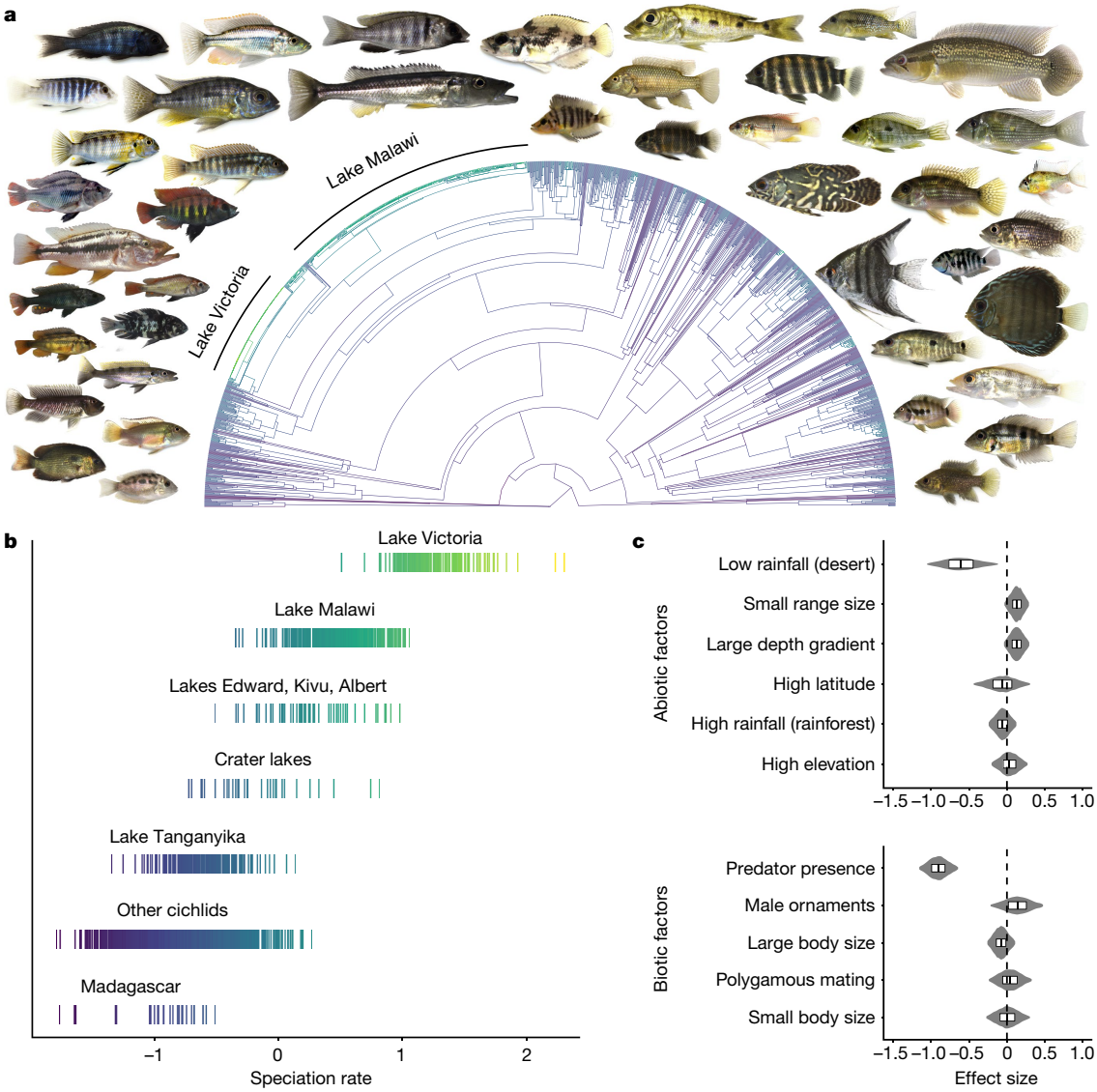


Fig. 1 | Cichlid speciation rates and their predictors on macroevolutionary timescales. a, A time-calibrated phylogeny of cichlids ($n = 1{,}712$ species), coloured by speciation rate, with the two fastest evolving clades, Lake Victoria and Lake Malawi, highlighted on the tree. Photographs show morphologies of cichlids that represent the vast phenotypic and ecological variation in this family. Branch colours correspond to the DR values shown in **b**. **b**, Distribution of speciation rates, measured using the DR statistic for all cichlid species. The highest speciation rates are found in Lake Victoria, Lake Malawi and the other lakes of the Lake Victoria Region Superflock—which comprises species of Lakes Edward, Albert, Kivu and species flocks in minor lakes of the region—followed by the crater lake flocks of Nicaragua and Cameroon, Lake Tanganyika, most cichlids and finally Madagascar cichlids. **c**, Effect size distributions of various extrinsic abiotic and extrinsic and intrinsic biotic predictors of speciation rate across all described cichlids. Box plots show the 50% credible intervals, violin plots show 95% credible intervals, and the middle line shows the posterior mean for each predictor. Photographs of fishes were produced by the authors (O.S. photographed the Lake Victoria cichlids, except *Prognathochromis macrognathus* (courtesy of F. Witte/HEST), M.D.M. imaged all other species).

vary by four orders of magnitude across the family, and that sustained explosive speciation rates are concentrated solely within several young lake radiations (Fig. 1a, b). These radiations, particularly that of Lake Victoria, possess some of the fastest sustained speciation rates observed in contemporary metazoans[15]. Note that the rates of these radiations are not only very high at the macroevolutionary scale, but that the extremely narrow and recent time window within which hundreds of speciation events occurred in a single lineage is also unparalleled in microevolutionary studies of speciation.

To examine the extent to which extrinsic environmental variables or general lineage traits influence variation in the speciation rate across the cichlid phylogeny, we assembled a comprehensive set of predictor variables. Both Bayesian regression analysis[16] and a nonparametric test[17] suggest that the presence of large visually oriented predatory fish[18] and an arid climate are major constraints to speciation rate in cichlids, whereas male-restricted ornamentation (indicative of evolution in response to sexual selection) and a wide gradient of water depth[19] show weak positive effects (Fig. 1c). However, when we fit hidden state speciation and extinction models[20], none of the environmental and trait factors provided a sufficient explanation for the uniquely high speciation rates that are observed in one of three lineages of haplochromine cichlids in Lake Victoria (Supplementary Information section 1.8). Further exploration with missing state speciation and extinction models[20] suggested three regimes for the speciation rate: a baseline rate, high speciation in Lake Malawi, Lake Xiloa (Nicaragua) and the Lake Victoria Region Superflock, and a uniquely extreme speciation rate in Lake Victoria itself. This suggests that our environmental and trait variables, albeit influential, are insufficient to explain the extreme

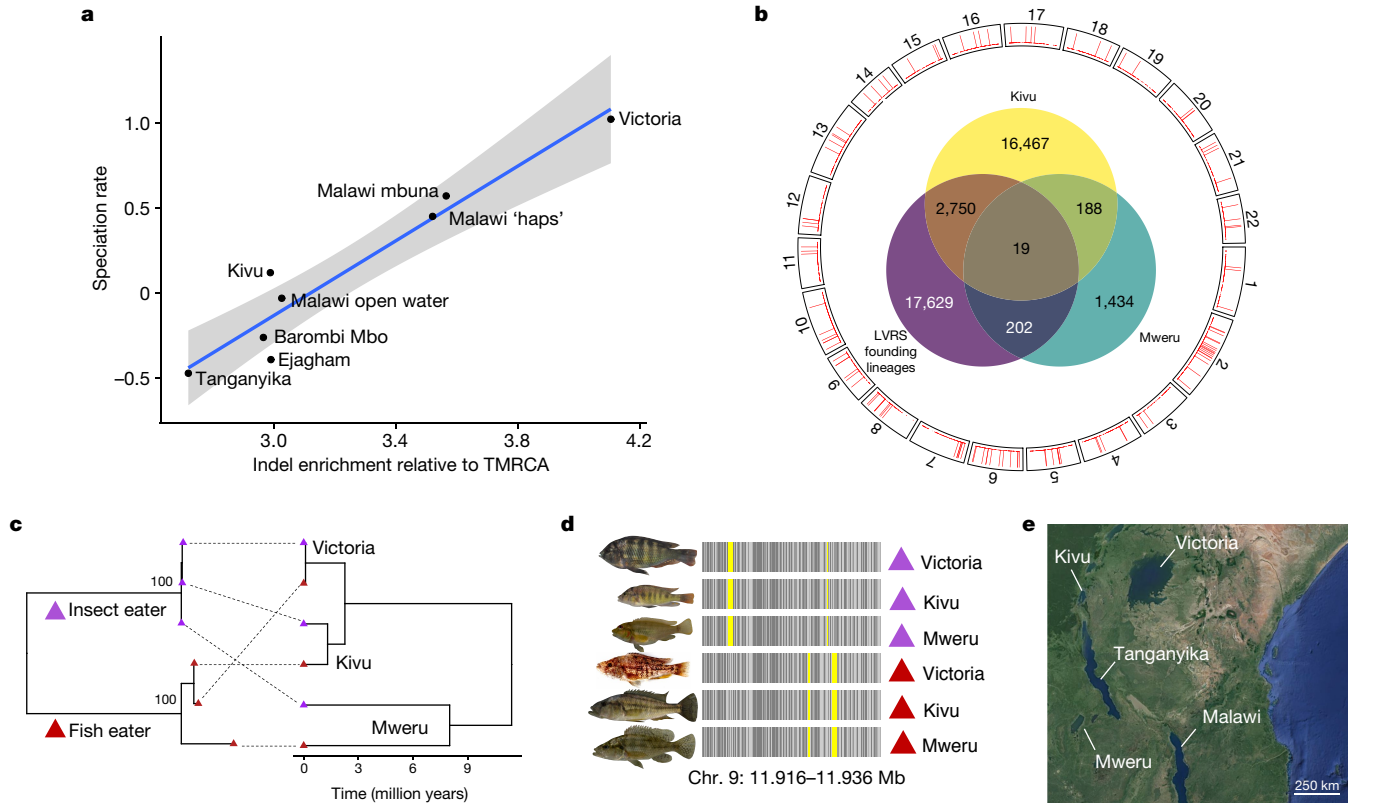


Fig. 2 | Large indels are associated with speciation and adaptation across cichlid radiations. a, Positive association between speciation rate (phylogenetic mean of the $\log_{10}$-transformed DR statistic; see Supplementary Information section 2.2 for additional speciation rate metrics) and the number of linkage-disequilibrium-filtered indels with a length of at least 5 bp relative to the TMRCA. The blue line indicates the regression line from a linear regression; grey shading shows the 95% confidence intervals of the mean. Indels are measured using genome pairs from Lake Tanganyika lamprologines, Barombi Mbo oreochromines, Ejagham *Coptodon*, Kivu haplochromines, Malawi open-water *Rhamphochromis*/*Diplotaxodon*/*Pallidochromis*, the ecologically diverse Malawi haplochromine clade known as 'haps', Malawi mbuna, and the Lake Victoria haplochromine species flock. **b**, Venn diagram of indels ($n = 38{,}689$) that segregate between species in the Lake Victoria radiation that also segregate either within Lake Kivu, Lake Victoria Region Superflock (LVRS) founding lineages and/or Lake Mweru. The outer ring shows the genomic positions of ancient large indels that were significantly associated with diet, habitat or nuptial coloration ($n = 214$). **c**, Phylogeny of the region on chromosome 9 that is deeply divergent between piscivores and other species in the Victoria radiation, with each allele having its closest relatives not within the Lake Victoria radiation but in morphologically and ecologically corresponding cichlids from Lake Kivu and Lake Mweru. The right phylogeny contains the same species from our cichlid megaphylogeny in Fig. 1. **d**, Multiple sequence alignment (20 kb) of the ancient chromosome 9 haplotype associated with piscivory shows multiple large indels (highlighted in yellow) that segregate across East Africa but are sorted into and fixed in a clade that contains all Lake Victoria piscivores. Species are, from top to bottom, *Paralabidochromis flavus* (Victoria), *Paralabidochromis paucidens* (Kivu), *Orthochromis* sp. 'red cheek' (Mweru), *Harpagochromis cavifrons* (Victoria), *Prognathochromis vittatus* (Kivu) and *Serranochromis* sp. 'checkerboard' (Mweru). All photographs were taken by O.S. **e**, Map showing the locations of Lakes Victoria, Kivu and Mweru, as well as Lake Tanganyika and the northern tip of Lake Malawi. Scale bar, 250 km. The map was generated using TerraMetrics (Google, 2020).

diversification rates of some clades, and that another process must be involved that is not fully captured by variables that represent variation in ecological opportunity and sexual selection alone.

We therefore examined the genomic substrate of adaptive radiation across two ecologically divergent species from adaptive radiations of each lake for which genomic data are available ($n = 8$ different radiations). Typically, speciation genomics is performed through alignment of small reads to a reference genome, but regions that contain insertions and/or deletions (indels) will not always align correctly. To avoid this issue, we assembled all genomes into de Bruijn graphs before reference alignment in order to accurately genotype indels[21]. We then counted the number of indels with a length of 5 base pairs (bp) or more that were fixed between the two species representing each radiation, filtering for linkage disequilibrium to account for the presence of multiple indels within large complex variants.

We find a positive association (Spearman correlation test, $P < 0.01$) (Fig. 2a and Supplementary Fig. 4) between speciation rate (estimated for each radiation from our large tree) and the number of divergently fixed indels relative to the time to the most-recent common ancestor (TMRCA). The fastest speciation rates by far are seen in Lake Victoria, and this radiation is enriched in a large number of indels that are divergently fixed between species even though the TMRCA is extremely short (Supplementary Information section 2.2). The lowest speciation rate, seen in a lineage of Tanganyikan lamprologines, was associated with many indels combined with a long TMRCA. Radiations with intermediate speciation rates either had a low number of indels fixed between the species and a recent TMRCA, as seen in Lake Ejagham, or a relatively high number of indels and a moderate TMRCA, as in the Malawi flocks. Our results suggest that the variation in speciation rate between these lineages in lakes is associated with whether a lineage has unusually many large indels for its age, highlighting the need to analyse indels as well as single-nucleotide polymorphisms in studies of adaptive radiation and speciation.

We next examined why Lake Victoria cichlids might contain more indels than expected for their relatively young age. We expanded our sampling of Victorian cichlids by producing assembly graphs of 100 whole genomes of 100 different species that represent every extant

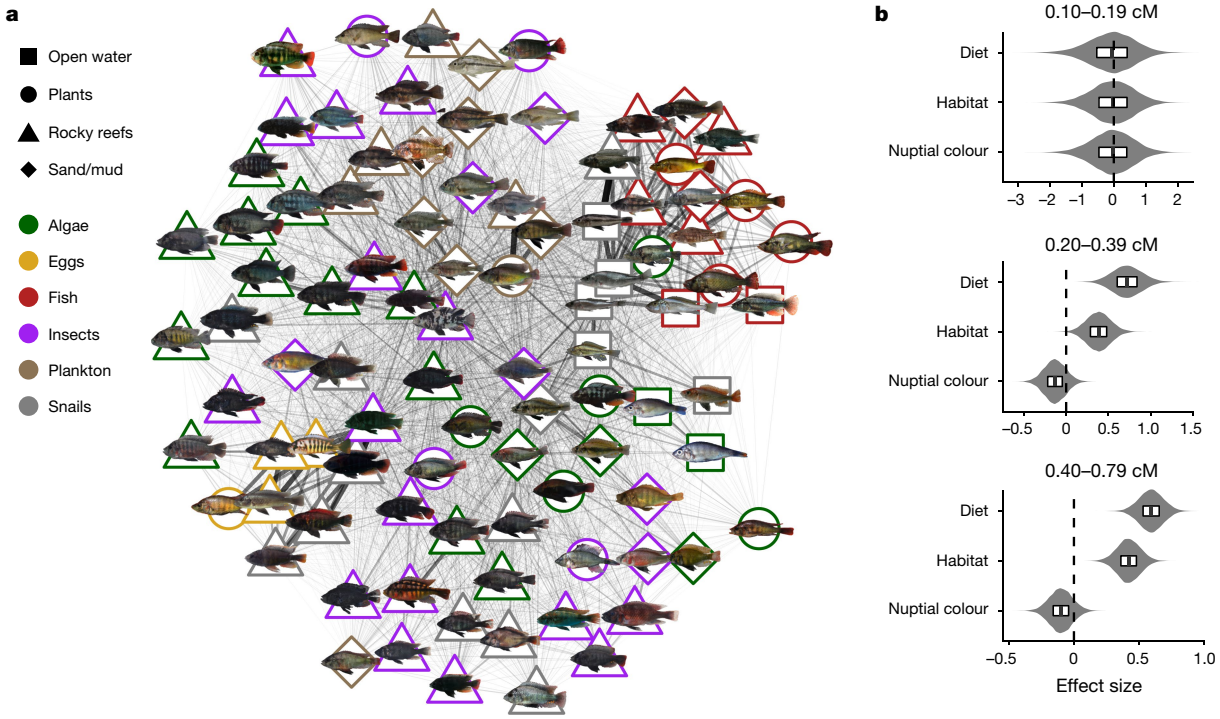


Fig. 3 | Networks of IBD blocks reveal the evolutionary history of the Lake Victoria cichlid radiation. a, IBD network of the Lake Victoria radiation. Nodes are the Lake Victoria cichlid species ($n = 92$) and links represent the number (thickness of links) of IBD segments ($n = 37{,}179$) shared between species. Species with similar diets and habitats exhibit more IBD sharing. **b**, Ecological predictors of the presence of IBD segments of a given size (ranging from 0.10 to 0.79 centiMorgan (cM)), taking into account the non-independence of pairwise comparisons ($n = 4{,}186$) between genomes. We use three different block size categories that represent different relative time points during the unfolding of the radiation and correspond to (1) coalescence of all species in a common ancestor; and (2) the beginning and (3) continuation of adaptive radiation into different ecological specialists with a structure driven by diet and habitat. Effect sizes (odds ratio = $e^{\text{effect size}}$) indicate whether IBD sharing between genomes is more likely between species with the same diet, habitat or male nuptial coloration than between species with different diet, habitat or male nuptial coloration. Box plots show the 50% credible intervals; violin plots show the full credible interval. Fish photographs were obtained by the authors (O.S. and M.D.M.).

taxonomic genus, dietary guild and habitat specialization in the radiation. We identified indels with a length of 5 bp or longer that segregate within the Lake Victoria radiation at a minor allele frequency of 0.01 or greater ($n = 251{,}690$ indels). We then examined whether these indels were also present in one or more outgroup species, which included haplochromine cichlids from the independent radiation in nearby Lake Kivu, cichlids that represent the two founding lineages of the entire Lake Victoria Region that includes both Victoria and Kivu (Congolese *Astatotilapia* and Upper Nile *Thoracochromis*), as well as representatives of two much older haplochromine genera (*Serranochromis* and *Orthochromis*) that diverged more than 10 million years ago from species of Lake Victoria and that are endemic to Lake Mweru in the upper Congo and represent the extreme ends of the ecomorphological gradient in Lake Mweru[22].

A large number of the indels that segregate between Lake Victoria cichlids also segregate between species across the outgroups (82,131 out of 251,690 indels; 43,442 fixed between outgroup taxa, 38,689 segregating within one or more outgroup taxa), suggesting that the Lake Victoria cichlid radiation contains many variants that existed up to 10 million years before the modern radiation (Fig. 2b). Genome-wide association analyses of these indels and a series of ecological variables that capture the major ecological dimensions of cichlid adaptive radiation revealed many indels that are associated with diet and habitat, and one indel that is associated with nuptial coloration, arranged in a highly polygenic architecture (Fig. 2b). Indel polymorphisms that segregate between species in Lake Victoria but not across our outgroups were much less often associated with these traits (Supplementary Information section 2.4). We discuss one of the ancient ecology-associated regions below.

On chromosome 9, we observe a region that is not only fixed across all morphologically specialized fish-eating Lake Victoria cichlid species that we sequenced (14 species) but is also seen in morphologically similar fish-eating specialists that are phylogenetically distant from the radiation. Several piscivory-associated indels in this region on chromosome 9 (Fig. 2c) occur in an intron of a gene that encodes an uncharacterized regulatory protein in the ubiquitin family that occurs adjacent to the mesenchymal stem cell gene *them6*. Fixation of this haplotype is shared between Lake Victoria piscivore species, the Lake Kivu predator *Prognathochromis vittatus* and predatory cichlids in the evolutionary distant genus *Serranochromis* in southern Africa (Fig. 2d). A comparison of internal branch lengths of gene trees that both flank and are within this region suggests that this is due to introgression rather than incomplete lineage sorting (Supplementary Information).

The evolutionary divergence in our large cichlid phylogeny suggests a most recent common ancestor between *Serranochromis* and the Lake Victoria Superflock roughly 10 million years ago (Fig. 2e), by far exceeding the divergence between lineages that gave rise to the hybrid swarm that founded the Lake Victoria Region Superflock. Currently the genus *Serranochromis* and the lineages ancestral to the Lake Victoria Superflock have very limited geographical range overlap (Fig. 2f); one rare species, *Serranochromis janus*, is known to occur in the nearby Malagarasi drainage[23] where it could have plausibly interacted with the founding lineages of the Lake Victoria Region Superflock, given that the region contains at least one extant representative of the Congolese founding lineage, *Astatotilapia stappersii*.

The presence of ancient indel polymorphisms, the hybrid origin[24] and the explosive speciation rate of the Lake Victoria cichlid radiation all pose substantial challenges for the reconstruction of the evolutionary

history of Lake Victoria cichlids, as evolution through reshuffling and recombination of ancient variants violates assumptions of a treelike evolutionary history, and incomplete lineage sorting is widespread. To overcome these challenges, we reconstruct networks from sharing of identity-by-descent (IBD) blocks[25,26] instead of using phylogenetic methods for the Lake Victoria radiation. IBD block sharing networks permit us to uncover how ecology and sexual selection interacted to enable the rapid speciation and colonization of many different adaptive zones and also permit us to reconstruct the relative timing of the emergence of the major ecological guilds.

We use the temporal information contained within IBD segment size[25] to examine the imprint of ecology and sexual selection on species differentiation at different stages of diversification in the Lake Victoria cichlid radiation. The sharing of small IBD segments represents older snapshots of the evolutionary relationships[27], whereas the sharing of larger segments indicates more-recent relationships. In all classes except the smallest IBD segment size class, which represents the earliest stages of the radiation, we find that both diet and macrohabitat have positive effects on species clustering in the network structure of the radiation such that more-closely related species share the same macrohabitat and belong to the same trophic guild significantly more often than less-closely related species.

We also found a smaller negative effect of male nuptial coloration on species clustering in the networks that reflect more recent ancestry (Fig. 3b), which indicates that species with recently shared ancestry are divergent in male nuptial coloration more often than expected and that genetically more-distantly related species more often share similar nuptial coloration (phylogenetic overdispersion of nuptial coloration in contrast to phylogenetic clustering of diet and habitat). These results suggest that young species—that is, shortly after speciation—are likely to exhibit different nuptial coloration, but similar ecologies, and is consistent with population-level studies on the mechanisms of speciation in the Victoria radiation[28]. Our IBD results are complemented by an analysis of *D*-statistics across the radiation that also found widespread non-treelike evolution as well as divergence driven by ecology and sexual selection (Extended Data Fig. 1).

The genomic architecture of adaptive radiation

Previous studies have proposed that cichlid adaptive radiations proceed in stages, in which macrohabitat divergence precedes dietary divergence[29], but this hypothesis is not supported for the Lake Victoria species flock. Instead, our results suggest that simultaneous divergence in diet and macrohabitat occurred and that this divergence depended critically on the exceptional genomic potential of these cichlids. This genomic potential is driven by an extraordinarily high density of ancient indel polymorphisms, many of which are directly linked to ecological divergence. Without this genetic variation, which consists of many ancient haplotypes that affect ecology and mating and could be recombined in many different ways to create the highly polygenic genomic architecture of this radiation, the number of new species that could have evolved and persist in sympatry in the 15,000 years since the filling of the lake would likely have been much more limited. Adaptive radiation across the food web would have been much slower, and very possibly never have occurred in the limited time, had these transitions been dependent on waiting times for new mutations.

Our results emphasize the need to use both large-scale comparative methods and fine-scale speciation genomics together to understand the drivers of speciation. Phylogenetic comparative methods are necessary to rigorously identify different evolutionary regimes, and—in particular—regimes with the high rates typical of some but not all adaptive radiations. Similarly, comparative genomics of many representatively chosen species within these radiations and their outgroups are required to discover and understand the non-treelike dynamics and genomically heterogenous ancestry that can characterize and drive these rapidly evolving regimes. Our insights into the factors that underlie cichlid diversification in general, and the special circumstances that permitted the exceptional rates of explosive adaptive radiation in Lake Victoria, make us hopeful that combining genomically informed speciation research with macroevolutionary analyses of diversification will soon enable a comprehensive understanding of why some clades produce spectacular radiations while others do not[30].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2652-7.

1. Otte, D. & Endler, J. *Speciation and its Consequences* (Sinauer, 1989).
2. Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer, 2004).
3. Gavrilets, S. *Fitness Landscapes and the Origin of Species (MPB-41)* Vol. 41 (Princeton Univ. Press, 2004).
4. Dieckmann, U., Doebeli, M., Metz, J. A. & Tautz, D. (eds) *Adaptive Speciation* (Cambridge Univ. Press, 2004).
5. Mayr, E. & Diamond, J. M. *The Birds of Northern Melanesia: Speciation, Ecology & Biogeography* (Oxford Univ. Press, 2001).
6. Nosil, P. *Ecological Speciation* (Oxford Univ. Press, 2012).
7. Seehausen, O. et al. Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).
8. Wolf, J. B. & Ellegren, H. Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* **18**, 87–100 (2017).
9. Flaxman, S. M., Wacholder, A. C., Feder, J. L. & Nosil, P. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* **23**, 4074–4088 (2014).
10. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
11. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
12. Witte, F. & Van Oijen, M. J. P. *Taxonomy, Ecology and Fishery of Lake Victoria Haplochromine Trophic Groups* (Nationaal Natuurhistorisch Museum, 1990).
13. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
14. Harvey, M. G. & Rabosky, D. L. Continuous traits and speciation rates: alternatives to state-dependent diversification models. *Methods Ecol. Evol.* **9**, 984–993 (2018).
15. Rabosky, D. L. et al. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395 (2018).
16. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
17. Rabosky, D. L. & Goldberg, E. E. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* **71**, 1432–1442 (2017).
18. Jackson, P. B. N. The impact of predation, especially by the tiger-fish (*Hydrocyon vittatus* Cast.) on African freshwater fishes. *Proc. Zool. Soc. Lond.* **136**, 603–622 (1961).
19. Wagner, C. E., Harmon, L. J. & Seehausen, O. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**, 366–369 (2012).
20. Beaulieu, J. M. & O'Meara, B. C. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* **65**, 583–601 (2016).
21. Turner, I., Garimella, K. V., Iqbal, Z. & McVean, G. Integrating long-range connectivity information into de Bruijn graphs. *Bioinformatics* **34**, 2556–2565 (2018).
22. Meier, J. I., et al. The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nat. Commun.* **10**, 5391 (2019).
23. Trewavas, E. *A Revision of the Genus* Serranochromis *Regan (Pisces, Cichlidae)* (Musée royal de l'Afrique Centrale, 1964).
24. Meier, J. I. et al. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).
25. Palamara, P. F., Lencz, T., Darvasi, A. & Pe'er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).
26. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
27. Ralph, P. & Coop, G. The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11**, e1001555 (2013).
28. Meier, J. I. et al. Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol. Ecol.* **26**, 123–141 (2017).
29. Streelman, J. T. & Danley, P. D. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* **18**, 126–131 (2003).
30. Gillespie, R. G. et al. Comparing adaptive radiations across space, time, and taxa. *J. Hered.* **111**, 1–20 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Genomic data are available at NCBI BioProject (PRJNA626405) other data are available from Dryad (https://doi.org/10.5061/dryad.fn2z34tr0), including trees and networks labelled by species.

Code availability

Scripts are available from Dryad (https://doi.org/10.5061/dryad.fn2z34tr0).

Acknowledgements We thank R. Bireley and the Sacramento Aquarium Society, P. George and the New England Cichlid Association, M. Welss and the IgV, R. Borstein of the Greater Chicago Cichlid Association and the American Cichlid Association for advice and access to specimens for photography in Fig. 1, K. Pedersen and J. Øgaard for Uganda fieldwork assistance along with NaFiRRI staff, including J. Balirwa, as well as cooperation from Kyoga region village elders, TAFIRI for hosting O.S. and team members during fieldwork, M. Kayeba, M. Haluna and H. Mrosso for assistance with Tanzanian fieldwork and K. Wagner, J. van Rijssel, F. Moser, O. Selz and L. Kaufman for discussions. This research was funded by Swiss National Science Foundation grant 31003A_163338 to O.S., L.E. and R.B., an American Cichlid Association Guy Jordan Fellowship to S.R.B. and Australian Research Council DE180101558 to M.D.M.

Author contributions M.D.M. and S.R.B. designed the phylogenetic analyses with assistance from O.S. and B.O'M. Trait data were gathered by S.R.B. and O.S. with assistance from M.D.M. GIS data were gathered by M.D.M. Genomic data was gathered by S.M., J.I.M. and M.D.M. M.D.M. and O.S. designed the genomic analyses with assistance from J.I.M. and advice from D.A.M., L.E. and R.B. M.D.M. performed fieldwork in Uganda with assistance from A.T. and coordination by O.S. O.S. and S.M. performed fieldwork in Tanzania with assistance from M.A.K. M.D.M. and O.S. wrote the manuscript with assistance from S.R.B., D.A.M. and J.I.M., with comments from and final version approval by all authors.

Competing interests The authors declare no competing interests.

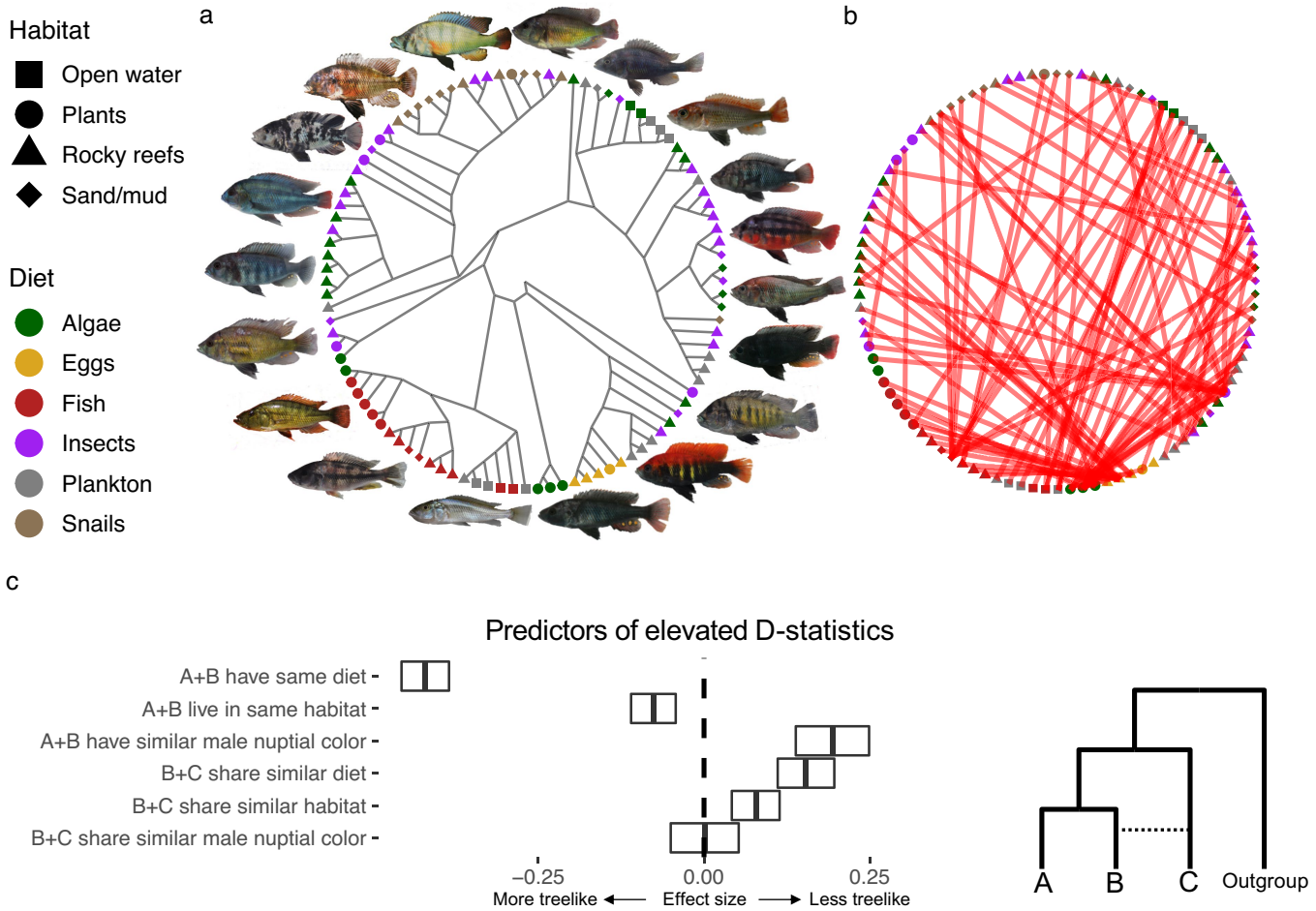
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2652-7.

Correspondence and requests for materials should be addressed to M.D.M. or O.S.

Peer review information Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | *D*-statistics across the Victoria radiation. a, Multispecies coalescent tree generated by ASTRAL for Lake Victoria cichlids ($n = 100$ species). Species diets and habitats are coded by colour and shape, respectively. **b**, The same arrangement of taxa as the coalescent tree, but highlighting the highest *D*-statistic associated with each genome. **c**, The 95% credible intervals of the ecology and sexual selection-based predictors of high introgression (Bonferroni-corrected $P < 0.05$ jackknife of the *D*-statistic) are shown, taking into account the non-independence of comparisons between the same genomes. Effect sizes (odds ratio = $e^{\text{effect size}}$) indicate whether introgression is more likely when the ingroup genomes (A + B) have a similar diet, habitat or male nuptial coloration, or whether introgression is more likely when the potentially introgressing genomes (B + C) do. Photographs were produced by the authors (O.S. and M.D.M.).

nature research

Corresponding author(s): Ole Seehausen

Last updated by author(s): Apr 26, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	R 3.6.1 PHLAWD (no version number listed) AnnotationBustR 1.2 RAxML-HPC v.8.12.2 phytools 0.6-99 rgbif 0.9.6 WorldClim 2 brms 2.10.0 STAN 2.19.1 ape 5.3 FISSE (no version number listed) hisse 1.9.5 dplyr 0.8.3 Dsuite 0.3 r21 SNPRelate 1.18.1 ASTRAL 5.14.3 McCortex 0.0.3-610 IBDseq r1206 PHLAWD 3.3a flexclust 1.4.0 TreePL 1.0 SplitsTree 4.15.1

Bali-Phy-3.4.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data and associated code is available on a Dryad digital repository at time of publication. Genome sequence data is available on the NCBI SRA at time of publication.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Compilation of published GenBank sequence data, data from geographic databases and the taxonomic literature, new whole genome sequencing of 100 species.
Research sample	Tissue samples from 100 Victorian cichlids from the tissue collection of OS.
Sampling strategy	One sample for each of 100 species that comprise the full range of ecological and taxonomic diversity in the radiation. Sampling was chosen on the basis of available funding in the project for genome sequencing, with a goal of maximizing ecological diversity for the given number of genomes able to be sequenced within the budget for the project.
Data collection	Whole genome sequencing, and collection via public repositories and the primary literature
Timing and spatial scale	We utilized preserved samples from the collection of OS, containing finclips from field-collected individuals from the years 1996-2014, with additional sampling performed by MM in Uganda in October 2015.
Data exclusions	No data was excluded from analyses.
Reproducibility	Citations of all data sources, and public deposition of new sequence data and curated sequence/trait/geographic data
Randomization	The work was not experimental in nature, and did not necessitate random assignment of individuals to treatment groups.
Blinding	No blinding

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Samples from Tanzania were from previous collection work by the Seehausen lab, spanning the years 1996-2014. In conjunction with NaFIRRI, Matthew McGee performed collections in Uganda during Ocrober 2015, prior to the onset of the rainy season.
Location	Ugandan cichlids were taken from the following locations: Lake Nawampasa (1.28, 33.35), Lake Gigati (1.29, 33.53), Lake Bisina (1.66, 33.86) and Banda Island (-0.25, 32.41) using gillnets and cast nets.
Access and import/export	Permits were obtained from the Ugandan Ministry of Agriculture and Fisheries with assistance from Uganda's National Fisheries Resource Research Institute. Samples were exported from Tanzania and Uganda in accordance with national laws and international guidelines.
Disturbance	Disturbance was minimized due to sampling a small number of individuals.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	MM took tissue samples and specimens from freshly killed cichlids caught by local fishermen using traditional methods overseen by NaFIRRI staff. The species, sex, and location of each individual is included in Appendix 5.
Field-collected samples	No laboratory work with live animals was performed.
Ethics oversight	All Uganda sampling was performed according to standard protocol from Uganda's National Fisheries Resource Research Institute.

Note that full information on the approval of the study protocol must also be provided in the manuscript.