

## Introduction



**Cite this article:** Sackton TB, Clark N. 2019 Convergent evolution in the genomics era: new insights and directions. *Phil. Trans. R. Soc. B* **374**: 20190102.

<http://dx.doi.org/10.1098/rstb.2019.0102>

Accepted: 18 April 2019

One contribution of 16 to a theme issue 'Convergent evolution in the genomics era: new insights and directions'.

**Subject Areas:**

evolution, genomics

**Keywords:**

convergent evolution, population genetics, comparative genomics

**Author for correspondence:**

Timothy B. Sackton

e-mail: [tsackton@g.harvard.edu](mailto:tsackton@g.harvard.edu)

Convergent evolution in the genomics era:  
new insights and directions

Timothy B. Sackton<sup>1</sup> and Nathan Clark<sup>2</sup>

<sup>1</sup>Informatics Group, Harvard University, Cambridge, MA, USA

<sup>2</sup>Computational and Systems Biology, University of Pittsburgh, PA, USA

TBS, 0000-0003-1673-9216

Convergent evolution—in which distinct lineages independently evolve similar traits—has fascinated evolutionary biologists for centuries [1], in large part because convergent evolution is often thought to represent a visible manifestation of the power of natural selection. Intuitively, what could explain convergent echolocation ability in bats and toothed whales except natural selection in response to foraging and hunting in environments with minimal visibility? Many other examples of phenotypic convergence—including limbless body plans in burrowing species [2], drug resistance in pathogens [3], and antifreeze proteins in arctic and antarctic fishes [4]—have similarly intuitive explanations as the result of adaptation to shared environments. Convergent evolution, then, can serve as a valuable proxy for repeated experiments in evolution, and understanding how convergent traits evolve, especially at the molecular level, has the potential to inform general rules about adaptation [5,6]. In the past decade, low-cost, high-throughput sequencing has ushered in an era of widespread genome sequencing in model and non-model organisms alike, bringing vast new data to bear on understanding convergence at the molecular level. In this special issue, we highlight the new insights that have emerged from applying the power of comparative and population genomics to the study of convergence, while also highlighting the challenges that still lie ahead.

A major impact of high-throughput sequencing has been an increasing focus on understanding the genetic underpinnings of convergent traits, to test whether and under what conditions phenotypic convergence is associated with convergence at the genetic level [7–10]. Convergence at the genetic level, of course, can reflect a variety of degrees of similarity, ranging from identical mutations fixed independently in different lineages, to evolutionary changes in the same genes but at different sites, to changes in the same pathways but at different genes. The observation of genetic convergence associated with a phenotypic trait can indicate that there are only a few feasible or easy genetic pathways to evolve a trait owing to constraints [5] or mutational biases [11], or could reflect instead a role of shared genetic variation [12]. Disentangling the role of these factors in adaptation has been a major motivation for studying the genomic basis of convergent traits.

The emerging trend from a variety of comparative and population genomic studies is that in many, but not all, cases, there is substantially more convergence at the genetic level among populations or species with convergent phenotypes than would be expected under null models [13–20]. In this issue, several studies extend these observations in new ways. Rennison *et al.* [21] analyse data from sticklebacks inhabiting many pairs of lake and stream environments, and using permutation tests show that peaks of local genetic differentiation overlap more often than would be predicted, implying genetic convergence in adaptation to aquatic environments. This observation of genetic convergence is also supported by Brown *et al.* [22], working in *Poecilia* species adapting to springs rich in hydrogen sulfide. In plants (*Arabidopsis* species) adapting to soils with toxic levels of heavy metals, Preite *et al.* [23] show modest evidence for genetic convergence and overlap in peaks of local adaptation between independent sites within species, but substantially less convergence between species. Rubin *et al.* [24] revisit the question of the molecular basis of independent transitions to eusociality in bees, and show that

while there is not strong evidence for convergent changes associated with origins of social behaviour, elaborations of eusocial complexity do appear to be associated with convergent changes in potential regulatory regions, highlighting the importance of considering the non-protein-coding portion of the genome in studies of convergence. Reviewing the evidence for convergence at the genetic level in domesticated cereal grasses both during domestication and afterward during adaptation to different crop environments; Woodhouse & Hufford [25] highlight both the extent of convergence (especially in domestication traits) but also the constraints imposed by domestication itself on later adaptation. Finally, Witt & Huerta-Sanchez [26] synthesize the evidence for convergent adaptation to high-altitude lifestyles in humans and domesticates, finding evidence for both convergent genetic responses and independent responses.

A second emerging trend is that those cases of convergent evolution with a shared genetic basis often involve selection acting on existing variation. Brown *et al.* [22] show that among independently evolving populations of fish (*Poecilia* sp.) living in environments rich in hydrogen sulfide, not only is there a substantial genomic signal for molecular convergence, but many selected alleles in independent populations share a common ancestry, implying convergence because of either introgression or selection on standing genetic variation. The role of introgression in facilitating convergent evolution is further discussed by Witt & Huerta-Sanchez [26], who highlight the observation that putatively adaptive alleles in key candidate genes (such as *EPAS1*) show repeated evidence for introgression. Distinguishing among modes of convergent evolution (selection on independent mutations, shared standing variation or introgressed alleles), and separating these from non-convergent evolution (e.g. an ancestral selective sweep) is a difficult challenge [27]. Lee & Coop [28] expand on these issues via a conceptual framework based on information theory, focused on quantifying the independent work done by natural selection in two (or more) populations or species as a way of thinking about how surprising (and thus interesting) the observation of shared genetic change in populations is.

While the focus of many studies of convergent evolution has a clearly adaptationist perspective, constraint can also be an important driver of patterns of convergence, and convergent evolution at the phenotypic or molecular level may not always reflect repeated selection to the same optimum [5]. If sources of variation are biased and only allow a limited number of changes, distantly related species may easily evolve convergent traits by non-adaptive processes. Such constraints are well known in the form of developmental constraint (e.g. canalization) and/or genetic constraints (e.g. deleterious pleiotropic effects). Mutational constraints also can play an important role in molecular convergence: Storz *et al.* [29] show that the high mutation rate at CpG dinucleotides leads to them preferentially contributing to convergent increases in oxygen affinity in haemoglobin in high-altitude bird species. As Yang *et al.* report [30], negative pleiotropy can also contribute to limiting the possible evolutionary paths to a new phenotype, and thus promoting convergence at the genetic level. They show that cardenolide resistance in Orthoptera is predictable based on results from other insect orders [31], in large part owing to the limited number of

possible substitutions in the  $\alpha$ -subunit of  $\text{Na}^+, \text{K}^+$ -ATPase that contribute to insensitivity without disrupting other phenotypes. In the other direction, epistasis and genetic background effects can limit molecular convergence across different species, as the phenotypic effects of specific mutations depend on the genetic background. A clear example of this is the lack of amino acid convergence in some species of high-altitude birds [16]. Finally, the relationship between convergent evolution and epistasis can be exploited to help understand genetic interactions: Fisher *et al.* [32] develop an approach using mutual information to identify mutations that co-occur more often than expected in different experimentally evolved populations. Such an approach promises to reveal genetic interactions that could not be detected using other experimental approaches.

In this issue, we also find contributions that explore in more depth existing challenges and potential ways forward for the study of convergent evolution. Several authors highlight areas where a deeper understanding of the phenotype in convergent phenotypes is important. Lamichhane *et al.* [33] argue for the importance of natural history knowledge in the study of the genomics of convergent phenotypes, and highlight the need for strong collaborations between organismal biologists and computational and population geneticists to unlock the full potential of the genomics era to elucidate the genetic basis of convergent traits. This theme is echoed by Fischer *et al.* [34] in the context of the convergent evolution of behaviour, where interdisciplinary studies between neuroscientists, evolutionary biologists and genome scientists are increasingly needed to untangle the extent to which convergent behaviours evolve with similar neurological or genetic mechanisms. Finally, the study by Song *et al.* [35] highlights the importance of thinking beyond single-organism phenotypes, by demonstrating some degree of convergence in the gut microbiome of distantly related blood-feeding vertebrates. As our understanding of the importance of the microbiome in host phenotypes grows, studies (e.g. [36]) looking at convergent evolution of biotic interactions across taxa will become increasingly important.

The rapid development of approaches to discover the genetic underpinnings of convergent evolution is energizing for the field, but we should remember to address potential confounders and unforeseen variables in our interpretation of apparent convergence. Multiple contributions in this issue point out, and begin to solve, some methodological challenges both arising from technical concerns and from interpretation. For one, Mendes *et al.* highlight the problem hemiplasy (which arises from discordance between gene trees and species trees) poses for studies of convergent evolution [37]. This discordance creates a tendency to make false inferences of molecular convergence, and they show that the rate of false positives can be alarmingly high. While simple remedies for this condition will not suffice, there are promising ways forward. Storz *et al.* [29] present evidence that a strong mutational bias leads to a bias in the types of convergent and adaptive substitutions that eventually fix between species. The implications are that investigators should be aware of how this and other unappreciated mutational biases could affect our inferential methods but also that there could be a general tendency of highly mutable sites to contribute disproportionately to adaptive changes. While several methods to infer convergently evolving amino acid

sites have been presented, another contribution from Rey *et al.* [38] contrasts those methods on the basis of their theoretical bases and practical performances. They raise the concern that contrasting definitions of molecular convergence can be confusing to the field and that inferential results often cannot be compared when differing definitions were used. Their study also evaluates the ability of different methods to infer amino acid convergence in the presence of non-adaptive background convergence, which occurs at a high rate simply because of the small number of possible molecular states. Finally, and crucially, Lee & Coop [28] remind us of the importance of not relying solely on patterns of convergence but that there must be inference of selection itself if we want to infer convergence owing to a shared selective pressure, rather than convergence owing to non-adaptive processes.

As evolutionary biologists, we have long been drawn to patterns of recurrent convergent change because of their ability to reveal evolutionary adaptations to life's challenges. Over the past decades, we have begun to discover the genetic mechanisms behind these convergent traits, and now, with increasingly easy access to full genome sequences, the rate of that discovery is rising rapidly. To mirror these exciting

developments, this special issue highlights recent progress in the field for the purposes of reflection and guidance for future studies. For one, comparative genomic studies of populations and species are proving to be powerful ways to answer long-standing questions about the process of adaptive evolution. Evidence is mounting that trait convergence is often achieved through genetic changes to shared pathways, genes or even molecular sites, and we may soon be able to describe rules that estimate the degree of genetic convergence expected for specific types of convergent traits. In counterpoint, while we naturally tend to focus on convergent adaptations to shared selective pressures, we should not overlook the numerous non-convergent genetic changes that allow species to adapt in their unique ways. Overall, the field is poised to provide answers to fundamental questions in evolutionary biology and genetics as studies of convergent evolution flourish again in the genomic era.

**Data accessibility.** This article has no additional data.

**Authors' contributions.** T.B.S. and N.C. wrote and edited the manuscript together.

**Competing interests.** We declare we have no competing interests.

**Funding.** N.C. was supported by NIH grant no. R01HG009299.

## References

- Darwin C. 1859 *On the origin of species*. London, UK: John Murray.
- Bergmann PJ, Morinaga G. 2019 The convergent evolution of snake-like forms by divergent evolutionary pathways in squamate reptiles. *Evolution* **73**, 481–496. (doi:10.1111/evo.13651)
- Farhat MR *et al.* 2013 Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **45**, 1183–1189. (doi:10.1038/ng.2747)
- Chen L, DeVries AL, Cheng CH. 1997 Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc. Natl Acad. Sci. USA* **94**, 3817–3822. (doi:10.1073/pnas.94.8.3817)
- Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840. (doi:10.1111/j.1558-5646.2011.01289.x)
- Orr HA. 2005 The probability of parallel evolution. *Evolution* **59**, 216–220. (doi:10.1111/j.0014-3820.2005.tb00907.x)
- Stern DL. 2013 The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764. (doi:10.1038/nrg3483)
- Stern DL, Orgogozo V. 2008 The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177. (doi:10.1111/j.1558-5646.2008.00450.x)
- Gompel N, Prud'homme B. 2009 The causes of repeated genetic evolution. *Dev. Biol.* **332**, 36–47. (doi:10.1016/j.ydbio.2009.04.040)
- Martin A, Orgogozo V. 2013 The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**, 1235–1250. (doi:10.1111/evo.12081)
- Chan YF *et al.* 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305. (doi:10.1126/science.1182213)
- Colosimo PF *et al.* 2005 Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307**, 1928–1933. (doi:10.1126/science.1107239)
- Sackton TB *et al.* 2019 Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78. (doi:10.1126/science.aat7244)
- Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2017 Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *Elife* **6**, e25884. (doi:10.7554/eLife.25884)
- Hecker N, Sharma V, Hiller M. 2019 Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. *Proc. Natl Acad. Sci. USA* **116**, 3036–3041. (doi:10.1073/pnas.1818504116)
- Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016 Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **354**, 336–339. (doi:10.1126/science.aaf9070)
- Kapheim KM *et al.* 2015 Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143. (doi:10.1126/science.aaa4788)
- Jones FC *et al.* 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61. (doi:10.1038/nature10944)
- Reid NM *et al.* 2016 The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354**, 1305–1308. (doi:10.1126/science.aah4993)
- Tishkoff SA *et al.* 2007 Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40. (doi:10.1038/ng1946)
- Rennison DJ, Stuart YE, Bolnick DI, Peichel CL. 2019 Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback. *Phil. Trans. R. Soc. B* **374**, 20180241. (doi:10.1098/rstb.2018.0241)
- Brown AP, McGowan KL, Schwarzkopf EJ, Greenway R, Rodriguez LA, Tobler M, Kelley JL. 2019 Local ancestry analysis reveals genomic convergence in extremophile fishes. *Phil. Trans. R. Soc. B* **374**, 20180240. (doi:10.1098/rstb.2018.0240)
- Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Krämer U, Yant L. 2019 Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. *Phil. Trans. R. Soc. B* **374**, 20180243. (doi:10.1098/rstb.2018.0243)
- Rubin BER, Jones BM, Hunt BG, Kocher SD. 2019 Rate variation in the evolution of non-coding DNA associated with social evolution in bees. *Phil. Trans. R. Soc. B* **374**, 20180247. (doi:10.1098/rstb.2018.0247)
- Woodhouse MR, Hufford MB. 2019 Parallelism and convergence in post-domestication adaptation in cereal grasses. *Phil. Trans. R. Soc. B* **374**, 20180245. (doi:10.1098/rstb.2018.0245)
- Witt KE, Huerta-Sánchez E. 2019 Convergent evolution in human and domesticate adaptation to high-altitude environments. *Phil. Trans. R. Soc. B* **374**, 20180235. (doi:10.1098/rstb.2018.0235)

27. Lee KM, Coop G. 2017 Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* **207**, 1591–1619. (doi:10.1534/genetics.117.300417)
28. Lee KM, Coop G. 2019 Population genomics perspectives on convergent adaptation. *Phil. Trans. R. Soc. B* **374**, 20180236. (doi:10.1098/rstb.2018.0236)
29. Storz JF, Natarajan C, Signore AV, Witt CC, McCandlish DM, Stoltzfus A. 2019 The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. *Phil. Trans. R. Soc. B* **374**, 20180238. (doi:10.1098/rstb.2018.0238)
30. Yang L, Ravikantachari N, Mariño-Pérez R, Deshmukh R, Wu M, Rosenstein A, Kunte K, Song H, Andolfatto P. 2019 Predictability in the evolution of Orthopteran cardenolide insensitivity. *Phil. Trans. R. Soc. B* **374**, 20180246. (doi:10.1098/rstb.2018.0246)
31. Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012 Parallel molecular evolution in an herbivore community. *Science* **337**, 1634–1637. (doi:10.1126/science.1226630)
32. Fisher KJ, Kryazhimskiy S, Lang GI. 2019 Detecting genetic interactions using parallel evolution in experimental populations. *Phil. Trans. R. Soc. B* **374**, 20180237. (doi:10.1098/rstb.2018.0237)
33. Lamichhaney S *et al.* 2019 Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. *Phil. Trans. R. Soc. B* **374**, 20180248. (doi:10.1098/rstb.2018.0248)
34. Fischer EK, Nowicki JP, O'Connell LA. 2019 Evolution of affiliation: patterns of convergence from genomes to behaviour. *Phil. Trans. R. Soc. B* **374**, 20180242. (doi:10.1098/rstb.2018.0242)
35. Song SJ *et al.* 2019 Is there convergence of gut microbes in blood-feeding vertebrates? *Phil. Trans. R. Soc. B* **374**, 20180249. (doi:10.1098/rstb.2018.0249)
36. Bittleston LS, Wolock CJ, Yahya BE, Chan XY, Chan KG, Pierce NE, Pringle A. 2018 Convergence between the microcosms of Southeast Asian and North American pitcher plants. *Elife* **7**, e36741. (doi:10.7554/eLife.36741)
37. Mendes FK, Livera AP, Hahn MW. 2019 The perils of intralocus recombination for inferences of molecular convergence. *Phil. Trans. R. Soc. B* **374**, 20180244. (doi:10.1098/rstb.2018.0244)
38. Rey C, Lanore V, Veber P, Guéguen L, Lartillot N, Sémon M, Boussau B. 2019 Detecting adaptive convergent amino acid evolution. *Phil. Trans. R. Soc. B* **374**, 20180234. (doi:10.1098/rstb.2018.0234)