

Evolutionary dynamics of recent selection on cognitive abilities

Sara E. Miller^a, Andrew W. Legan^a, Michael T. Henshaw^b, Katherine L. Ostevik^c, Kieran Samuk^c, Floria M. K. Uy^{a,d}, and Michael J. Sheehan^{a,1}

^aDepartment of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853; ^bDepartment of Biology, Grand Valley State University, Allendale, MI 49401; ^cDepartment of Biology, Duke University, Durham, NC 27708; and ^dDepartment of Biology, University of Miami, Coral Gables, FL 33146

Edited by Gene E. Robinson, University of Illinois at Urbana–Champaign, Urbana, IL, and approved December 29, 2019 (received for review November 2, 2019)

Cognitive abilities can vary dramatically among species. The relative importance of social and ecological challenges in shaping cognitive evolution has been the subject of a long-running and recently renewed debate, but little work has sought to understand the selective dynamics underlying the evolution of cognitive abilities. Here, we investigate recent selection related to cognition in the paper wasp *Polistes fuscatus*—a wasp that has uniquely evolved visual individual recognition abilities. We generate high quality de novo genome assemblies and population genomic resources for multiple species of paper wasps and use a population genomic framework to interrogate the probable mode and tempo of cognitive evolution. Recent, strong, hard selective sweeps in *P. fuscatus* contain loci annotated with functions in long-term memory formation, mushroom body development, and visual processing, traits which have recently evolved in association with individual recognition. The homologous pathways are not under selection in closely related wasps that lack individual recognition. Indeed, the prevalence of candidate cognition loci within the strongest selective sweeps suggests that the evolution of cognitive abilities has been among the strongest selection pressures in *P. fuscatus*' recent evolutionary history. Detailed analyses of selective sweeps containing candidate cognition loci reveal multiple cases of hard selective sweeps within the last few thousand years on de novo mutations, mainly in noncoding regions. These data provide unprecedented insight into some of the processes by which cognition evolves.

cognition | selective sweeps | evolutionary dynamics | wasps | social evolution

Cognition is arguably among the most complex animal traits and has been instrumental in the ecological and evolutionary success of disparate lineages (1). Decades of research have documented patterns of cognitive variation among animals, consistent with environmental and social selective forces shaping cognitive abilities (2, 3). Much research on cognitive evolution has focused on brain size as a proxy for cognitive abilities and utilized comparative methods to identify life-history, ecological, and social factors that explain variation in brain investment (4–6). Many studies, primarily in birds and mammals, have sought to document the ecological and evolutionary consequences of larger brains (7, 8). Increasingly, researchers are tackling the issue of how cognition evolves by measuring the heritability of cognitive abilities in wild populations (9), attempting to link cognition to fitness (10) and using experimental evolution (11). There are also attempts to understand the genetic architecture of variation in cognition. For example, a genome-wide association study of educational attainment in humans identified hundreds of loci that jointly explained ~10% of variance in cognitive performance (12). The presence of standing genetic variation in cognitive traits has been found in many species, suggesting that selection on standing variation can produce substantial shifts in cognitive abilities. This intuition is consistent with changes in cognitive abilities seen during the recent history of animal domestication and in experimental evolution studies (11, 13–15). Collectively, these studies indicate that, within species, variation

in cognitive abilities is a heritable quantitative trait (9, 16). Although these approaches have generated many hypotheses for the selective forces driving the evolution of novel cognitive abilities, these studies do not provide direct evidence for the evolutionary dynamics underlying cognitive evolution and, as a result, the process by which cognitive abilities evolve remains an essentially unexplored question.

Analyses of genomic patterns of selection have the potential to reveal the mode and tempo of cognitive evolution. Prior comparative genomic analyses have identified signatures of positive selection on genes associated with the brain or nervous system in a number of species (17–19). While these data show that neural systems can be shaped by positive selection, the deep timescale of such analyses obscures the evolutionary dynamics. Population genomic scans have detected evidence of recent selection on genes annotated with functions in the nervous system and cognition in diverse species (20–23). Together, these studies provide molecular evidence that cognition is shaped by natural selection although the nature and timing of the selective events have not been analyzed. The ability to detect the mode and tempo of selection decays with time (24) so an ideal system for examining the process of cognitive evolution requires a species that has evolved cognitive abilities absent in close relatives, indicating

Significance

Humans have long been fascinated by animal cognition, but little research has addressed the dynamics of the evolution of cognitive traits. Does cognitive evolution involve strong selection on novel mutations or selection on preexisting genetic variation? Do selective events happen in concert or in multiple independent bouts? We examined population genomic signatures of cognitive adaptation in *Polistes fuscatus* paper wasps, which have recently evolved individual facial recognition. We find evidence for multiple hard selective sweeps of novel mutations associated with genes involved in learning, memory, brain development, and visual processing. Arguably, selection on cognition has been among the strongest selective pressures in the species' recent history. These data provide insight into the evolutionary processes by which new cognitive traits evolve.

Author contributions: S.E.M. and M.J.S. designed research; S.E.M., A.W.L., M.T.H., K.L.O., K.S., F.M.K.U., and M.J.S. performed research; S.E.M. and M.J.S. analyzed data; and S.E.M. and M.J.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: Sequence data are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive, <https://www.ncbi.nlm.nih.gov/sra> (Bioproject accession no. PRJNA482994).

¹To whom correspondence may be addressed. Email: msheehan@cornell.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1918592117/-DCSupplemental>.

recent cognitive evolution. From this starting point, we can then interrogate the magnitude and nature of the selective signature on loci related to cognition and brain function in a focal species compared to its relatives. Such a system would provide novel insights into the mode and tempo of cognitive evolution.

The evolution of visual individual recognition in the northern paper wasp (25), *Polistes fuscatus*, provides an unusual opportunity to examine the mode and tempo of cognitive evolution. *P. fuscatus* has uniquely evolved the ability to learn and remember conspecific facial images (26). Facial processing in *P. fuscatus* is likely not the result of an overall increase in cognitive ability as other species of paper wasp are equally adept at learning nonface images (26), but rather a novel cognitive trait related to recognizing and remembering other individuals. Diverse facial patterns mediate individual recognition in *P. fuscatus*, but they are absent in close relatives that lack visual individual recognition (25) (Fig. 1B). In *P. fuscatus*, nests are initiated each spring by one or more mated females (27). Individual recognition mediates dominance interactions among nest cofoundresses (25) and has been associated with complex social cognition in *P. fuscatus*, including highly robust social memories (28) and tracking others' contribution to work and egg laying (29). Multiple paper wasp species cooccur throughout much of the range of *P. fuscatus*, often nesting on the same buildings and capturing the same prey species (27). Thus, differences in facial recognition abilities represent a prominent difference between *P. fuscatus* and its relatives.

Taking advantage of the paper wasp system, we seek to address four interrelated questions regarding the mode and tempo of cognitive evolution.

1) What is the population genomic signature of cognitive evolution? Depending on the mode and tempo of evolution, the population genomic signature of evolutionary changes in

phenotypes will vary (30, 31). Here, we consider two scenarios. First, the evolution of cognitive traits may be highly polygenic (Fig. 1A), with the total effect of selection spread across many loci with small effects on fitness (32). A highly polygenic model predicts selection on any given cognition allele is likely to be relatively weak, although potentially detectable in aggregate. Second, the evolution of cognitive abilities may involve stronger selection on a more modest number of loci (Fig. 1A) (31, 33). Under this oligogenic model, we would expect to observe particularly strong signatures of selection on a more limited set of loci annotated with cognition-related functions. Thus, determining whether or not there is a detectable signature of selection—and the relative strength of evidence for selection among relevant loci—allows for inferences regarding the recency, strength, and genomic architecture of selection on cognition for a particular lineage.

2) Is there evidence of hard sweeps on de novo mutations in cognition annotated loci? Adaptive evolution may proceed by selection acting on standing variation or de novo mutations (Fig. 1A) (30, 34, 35). Selection acting on standing genetic variation (i.e., alleles present in multiple copies in a population) can lead to soft selective sweeps, partially reducing genetic diversity surrounding the selected site by increasing the frequencies of multiple haplotypes containing the target mutation (36). Soft sweeps of standing variation provide the potential for rapid evolution in the face of novel selection pressures (35). The abundant evidence for heritable variation in cognitive abilities in wild populations (9) suggests that selection on standing variation is likely to be involved in cognitive adaptation. Adaptive change may also occur through hard selection on de novo mutations, which produces a genetic signature of a hard selective sweep characterized by a

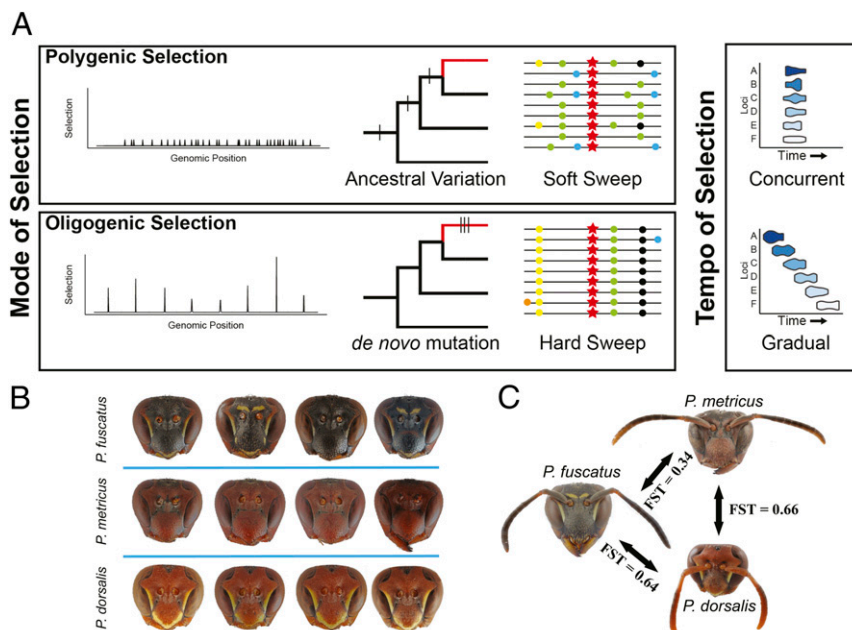


Fig. 1. (A) Graphical representation of two potential hypotheses for the mode and tempo of cognitive evolution. Polygenic selection on many loci from multiple segregating alleles of small effect would result in many soft selective sweeps whereas oligogenic selection from de novo mutations would be detectable as hard selective sweeps across a moderate number of loci. The cartoon graphs depict the predicted patterns for each hypothesis. The leftmost cartoons shows selective sweeps, with genomic position on the x-axis and strength of selection on the y-axis. Hatch marks on the phylogenetic tree indicate the branch where a selected mutation arose. Simplified haplotypes given to illustrate soft versus hard sweeps. Haplotypes are depicted with polymorphisms shown as different colored shapes. A change in the selective landscape could lead to concurrent selection on many loci; alternatively, a mutation order process could lead to selection on different loci over time. The cartoons illustrate the estimate of the relative timing of selection for loci A–F, shown as horizontal violin plots. (B) *P. fuscatus* has greater phenotypic variation in facial coloration than other species of *Polistes*, as exemplified by variation in female facial coloration within a single nest for *P. fuscatus*, *P. metricus*, and *P. dorsalis*. Faces were photographed with antennae removed. (C) Pairwise F_{ST} among *P. fuscatus*, *P. metricus*, and *P. dorsalis* show limited genetic differentiation between species.

- broader region of reduced genetic diversity with a single dominant haplotype (36). Signatures of hard sweeps of de novo mutations are most detectable when selection is relatively strong and recent (37). Moreover, hard sweeps of de novo mutations would suggest that standing variation may not be fully sufficient to generate the observed shifts in cognitive abilities and that cognitive evolution is constrained by the availability of adaptive mutations (30). Thus, evidence of hard selective sweeps from de novo mutations linked to loci annotated with cognition-related functions in *P. fuscatus* would provide novel insights into the relative mutational constraints, as well as the strength of selection mediating cognitive evolution.
- 3) What is the relative evidence for selection on coding versus regulatory evolution related to cognitive evolution? Two prominent mechanisms through which mutations can influence phenotypes and be exposed to selection are by either changing the coding sequence of a gene, thereby affecting the protein product, or by changing regulatory elements that influence the relative levels, timing, locations, or splice forms of that gene (38). Whereas coding sequence changes can affect the structure of a protein, regulatory mutations can act in specific tissues or developmental stages (39). Many studies looking for signatures of selection in relation to cognition have examined protein-coding evolution (17–19) although there is evidence of selection on regulatory elements with suggested links to cognition as well (19). More generally, there is a broad consensus of the importance of regulatory evolution in shaping patterns of phenotypic change (40). The relative signature of selection on coding versus regulatory mutations remains an open question with regard to cognitive evolution.
 - 4) What is the relative timing of selective events? Identifying the relative ages of selected alleles will reveal the tempo of cognitive adaptation (41). The tempo of adaptive evolution is a function of the strength of selection, as well as the timing of adaptive mutations. For a single adaptive allele, the age of the allele provides information about the speed of evolution. Comparing the relative ages of multiple adaptive alleles can help test hypotheses regarding possible mechanisms driving change in cognitive ability (42). For example, a change in ecological or social pressures modifying the selective landscape could lead to concurrent selection on many loci until the population reaches a new fitness optimum (43). Alternatively, the presence of multiple discreet bouts of selection may imply that evolving novel cognitive abilities is limited by the availability of adaptive mutations (Fig. 1A) (44). Little work has set out to explicitly address the selective dynamics of cognitive evolution, and insights from this process have important implications for our understanding of how novel complex traits have evolved across animal lineages.

De Novo Assembly of Three High Quality *Polistes* Wasp Genomes

We explored the genomic signatures of recent evolution by first assembling and annotating high quality de novo genomes for *P. fuscatus* and two closely related species lacking facial recognition, *Polistes metricus* and *Polistes dorsalis* (Fig. 1B). The three de novo genomes are ~220 megabases (Mb), among the most highly contiguous Hymenopteran genomes to date, and contain a nearly complete set of highly conserved genes (Table 1). Genomic features were comparable to previously sequenced paper wasp genomes (*SI Appendix*, Figs. S1–S5 and Tables S1 and S2, and refs. 45 and 46). These species provide an unusually tractable system for identifying the targets of recent natural selection. Pairwise fixation index (F_{ST}) calculations (Fig. 1C and *SI Appendix*, Fig. S6) show only moderate genetic differentiation among species, in effect limiting the genomic regions in *P. fuscatus* potentially associated with recent cognitive evolution in this lineage. Analysis of population samples shows rapid decay of linkage disequilibrium (*SI*

Table 1. Summary statistics for *P. fuscatus*, *P. metricus*, and *P. dorsalis* de novo genome assemblies

	<i>P. fuscatus</i>	<i>P. metricus</i>	<i>P. dorsalis</i>
Genome Assembly			
Scaffolds > 5,000 bp	187	216	1,075
Scaffolds > 50,000 bp	185	151	108
Scaffold N50	9.11 Mb	4.63 Mb	5.37 Mb
Scaffold N75	4.45 Mb	2.7 Mb	2.39 Mb
Scaffold L50	8	14	13
Scaffold L75	16	29	27
Largest scaffold	19.6 Mb	16.0 Mb	20.3 Mb
Total length	219 Mb	219 Mb	209 Mb
GC content, %	32.74%	31.61%	32.54%
No. of genes	12,034	12,032	11,311
Genome size			
In silico estimate	227 Mb	205 Mb	261 Mb
BUSCO			
Complete BUSCOs (eukaryote)	296/303	297/303	297/303
Complete BUSCOs (arthropod)	1,060/1,066	1,066/1,066	1,064/1,066

GC, guanine and cytosine.

Appendix, Fig. S7), consistent with high recombination rates reported in other social hymenopterans (47). This combination of features makes it possible to identify regions under selection with exceptionally high precision.

Recent Strong Selection on Visual Cognition, Neural, and Learning and Memory Annotated Loci in *P. fuscatus*

We conducted a genome-wide scan for selective sweeps in *P. fuscatus* using 40 resequenced genomes from two populations (mean coverage 12.2×) (*SI Appendix*, Table S3). Selective sweeps were determined with SweepFinder2 (48), which calculates a composite likelihood ratio (CLR) based on shifts in the allele frequency spectrum consistent with selection. By including individuals sampled from two different populations, selective sweeps should be detected only in genomic regions that are under selection in both populations and not in genomic regions that are under selection as a result of local adaptation. Using simulations, we first demonstrated that CLR values increase with the strength and recency of selection and that CLR values are greater when selection acts upon new mutations rather than standing variation (*SI Appendix*, Figs. S8 and S9). Highly polygenic selection would thus be detectable as a minor positive shift in CLR values for target gene sets associated with cognition, compared to the genome-wide background, whereas oligogenic selection would cause greatly elevated CLR across a moderate number of loci with cognition-related annotations (Fig. 1A).

The genome scan identified many narrow selective sweeps spread across the genome. We first examined the highest CLR peaks consistent with recent, strong selective sweeps using a stringent P value of $<5e^{-8}$ based on the Z-distribution of CLR scores (Fig. 24). This cutoff identified 138 selective sweeps that were narrow (median 2,843 base pairs [bp]; range 100 to 56,811 bp) and typically coincided with regions of elevated genomic differentiation from a reconstructed ancestral genome (Fig. 24 and *SI Appendix*, Fig. S10), lending further support of recent lineage-specific evolution at these genomic regions. Many of the selective sweeps encompassed annotated gene boundaries, but others fell nearby annotated genes but did not overlap with coding regions or introns, suggesting a potential for regulatory evolution. We identified 183 genes within 5 kilobases (kb) of the top sweeps that are potential targets of selection (median = 1 gene per sweep, range: 0 to 8 genes). Of the 138 selective sweeps, 39 (28%) were associated with loci implicated in learning,

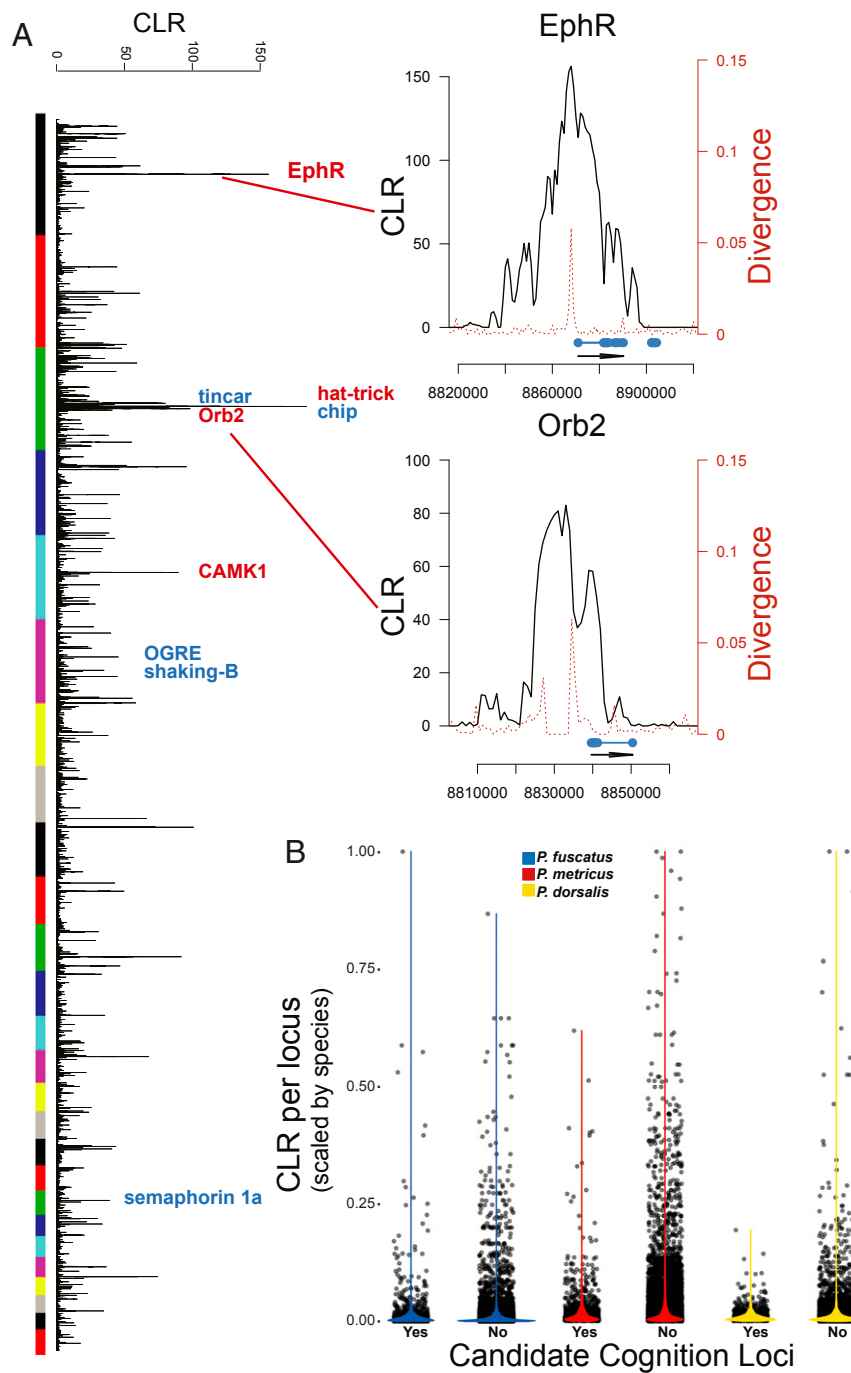


Fig. 2. (A) Narrow peaks of elevated composite likelihood ratio (CLR) values indicate genomic regions in *P. fuscatus* characteristic of recent selective sweeps. CLR peaks occur throughout the *P. fuscatus* genome. Peaks frequently coincide with regions of increased divergence from the *P. fuscatus* and *P. metricus* ancestral genome. A subset of candidate cognition- and visual processing-related genes located within CLR peaks annotated with GO terms related to learning and memory (red) and compound eye development (blue) are indicated. CLR and divergence values were calculated in 1,000-bp windows. Colored bars show the position of each scaffold (first 26 shown). A detailed view of two selective sweeps shows CLR peaks located near genes (blue bars) involved in insect memory formation and visual processing, *EphR* and *Orb2*. Divergence along the *P. fuscatus* lineage from a reconstructed genome of the most recent common ancestor of *P. fuscatus* and *P. metricus* is indicated with the dotted red line. The narrow sweeps shown here are among the widest in *P. fuscatus*, emphasizing the precision of localizing selection in this species. (B) CLR values for potential candidate cognition loci annotated with visual learning and cognition related functions ("Yes") are greater in *P. fuscatus* than in *P. metricus* or *P. dorsalis*. All other genes lacking GO annotation associated with visual processing, cognition, brain development, or learning and memory are labeled as "No". CLR values have been scaled to compare across the three species.

memory, neurogenesis, or the insect visual system based on gene ontology (GO) terms of the gene within the locus (Dataset S1). Notable candidate loci include (but are not limited to) ephrin receptor tyrosine kinase receptor (*EphR*), which is important for

brain development (49) and has been experimentally shown to mediate learning and memory in honey bees (50); *Orb2*, which plays a key role in the formation of long-term memories (51); and optic ganglion reduced (*OGRE*), which is expressed in the

optic lamina of developing *Drosophila* and is required for photoreceptor development (52).

The strongest sweeps in *P. fuscatus* were significantly enriched for GO terms related to neuronal development and synapse organization ($P < 0.0001$) (SI Appendix, Table S4), matching our prediction that recent cognitive evolution in *P. fuscatus* is associated with evolution in genes annotated with cognition and nervous system GO terms. Similarly, an analysis of all genic loci (an annotated gene body $\pm 5,000$ bp) ranked by top CLR score within each locus reveals significant enrichment of GO terms related to vision, brain development, and cognition among the loci with the strongest evidence of recent selection (Dataset S1 and SI Appendix, Fig. S11). This pattern of particularly strong selection on a moderate number of loci is inconsistent with a strictly polygenic model of cognitive adaptation, which predicts weaker signatures of selection spread across numerous loci. Instead, the enrichment of cognition annotated loci within the most extreme CLR peaks fits an oligogenic model of cognitive adaptation.

Recent Extreme Selection on Cognition Loci Is Specific to *P. fuscatus*

We next asked whether the strong signature of recent selection on cognitive abilities was specific to *P. fuscatus*, which has notably evolved visual individual recognition, or if other related species of paper wasps also show recent selection on cognition related loci (Fig. 1B and C). If similar patterns of selection are shown in all species, that would suggest that cognitive evolution in *P. fuscatus* is not associated with a distinct selective signature. In contrast, if *P. fuscatus* shows a distinct signature of selection on cognition annotated loci, it would suggest the patterns are related to recent lineage-specific cognitive evolution in *P. fuscatus*. We conducted selection scans in two closely related species for which we had generated de novo genomes, *P. metricus* and *P. dorsalis* (Fig. 1B and SI Appendix, Fig. S12), that lack individual recognition. First, we observed that orthologs of the highly selected loci in *P. fuscatus* showed no evidence of recent selection in either *P. metricus* or *P. dorsalis*. Instead, we find that loci associated with recent, strong selection in *P. dorsalis* are retrotransposases and in *P. metricus* loci associated with olfactory receptor activity and hydrocarbon production. Next, we tested the hypothesis that loci potentially associated with cognition were under stronger, harder, and/or more recent selection in *P. fuscatus* relative to the other species. Based on the cognition-related phenotypes known to have evolved in *P. fuscatus* in association with individual recognition (25, 53, 54), loci with genes annotated with gene ontology (GO) terms for cognition, mushroom body development, visual behavior, learning or memory, and eye development were classified as “candidate cognition loci.” Given that we would not expect strong selection to occur in all loci annotated with these functions, this test is conservative as it should be inherently biased against detecting weak evidence for selection in a large dataset. Across all species, CLR values for potential candidate cognition loci showed a strong interaction between locus type and recognition system (Fig. 2B) (locus type: $F_{1,34361} = 0.72$, $P = 0.40$; recognition system: $F_{1,1} = 0.42$, $P = 0.63$; locus type \times recognition system: $F_{1,34361} = 24.9$, $P < 0.0001$). Thus, evidence for positive selection on genic loci annotated as visual perception, cognition, memory, and neural development is far stronger in *P. fuscatus* than in either *P. metricus* or *P. dorsalis*. Similar results were observed when considering specific GO term annotations related to cognition (e.g., mushroom body development: locus type: $F_{1,34361} = 0.72$, $P = 0.40$; recognition system: $F_{1,1} = 0.61$, $P = 0.58$; locus type \times recognition system: $F_{1,34361} = 10.9$, $P = 0.0009$; eye development: locus type: $F_{1,34361} = 0.27$, $P = 0.6$; recognition system: $F_{1,1} = 0.61$, $P = 0.58$; locus type \times recognition system: $F_{1,34361} = 8.73$, $P = 0.003$) (SI Appendix, Fig. S13). The different pattern of selection on cognition-related GO terms in *P. fuscatus* is not

caused by a general difference in the ability to detect selection among species as loci that would be predicted to be under selection in all species show similar patterns of selection. For example, loci annotated with immune response show similarly elevated CLR values across all three species (locus type: $F_{1,34361} = 5.08$, $P = 0.024$; recognition system: $F_{1,1} = 0.60$, $P = 0.58$; locus type \times recognition system: $F_{1,34361} = 0.35$, $P = 0.55$) (SI Appendix, Fig. S13). Traditional GO term analysis of the outlier sweeps ($P < 5e^{-8}$ based on Zscore of CLR values) in the other species (SI Appendix, Table S4) and a gene set enrichment analysis (Dataset S1) failed to find evidence of strong selection on GO terms related to cognition in either *P. metricus* or *P. dorsalis*. Thus, *P. fuscatus* shows a distinct pattern of recent, strong selection on loci related to vision, learning and memory.

The unique pattern of selection on candidate cognition loci in *P. fuscatus* is consistent with selection on cognitive abilities underlying individual recognition. Although individual recognition is unlikely to have been the only selective pressure in the *P. fuscatus* lineage, it is an obvious and prominent difference in cognitive abilities between *P. fuscatus* and its relatives (26, 55). Individual recognition appears to provide a parsimonious explanation for the observed differential patterns of selection on cognition loci among the three species examined. Future work is needed to determine precisely whether and how particular selected mutations contribute to individual recognition abilities. Nevertheless, the evidence of multiple strong selective sweeps, including loci annotated with functions related to learning, memory, visual perception, neural development, and cognition, suggests that selection on cognitive traits, likely driven by individual recognition, has been among the strongest selection pressures in recent history in the *P. fuscatus* lineage.

Evidence for Multiple Hard Sweeps of De Novo Mutations Associated with Cognitive Evolution

The speed and dynamics of selection are expected to differ between traits evolving from standing variation or de novo mutations (56, 57). To determine which alleles are unique to *P. fuscatus*, we generated population-scale resequencing data for 93 individuals from four closely related species—including resequencing data for two additional close relatives, *Polistes carolina* and *Polistes perplexus*, in addition to the previously generated samples for *P. metricus* and *P. dorsalis* (SI Appendix, Fig. S14 and Tables S3 and S5). For 4,880,035 single-nucleotide polymorphisms (SNPs), we were able to classify variants by comparing the allele frequency in *P. fuscatus* with the allele frequency in these related species. We classified sites as fixed de novo mutations when *P. fuscatus* was fixed for an allele that was not present in other species, and as segregating de novo polymorphisms when sites were fixed in all other species, but variable in *P. fuscatus*. Fixed ancestral polymorphisms were sites fixed in *P. fuscatus* but for which a different allele is present in at least two other species. Shared polymorphisms are those sites that are segregating in *P. fuscatus* and at least one other species. Sweep regions in *P. fuscatus* were greatly enriched for fixed de novo mutations and fixed ancestral polymorphisms relative to the rest of the genome (SI Appendix, Table S5) (McNemar’s test, $P < 0.0001$), providing independent evidence of selection in these regions. Based upon these classifications, 46 (33.3%) of the top sweeps contained at least one fixed or nearly fixed de novo mutation, consistent with hard selective sweeps. We secondarily used a machine learning approach to classify regions as hard or soft sweeps (58). The two methods produced largely corresponding results, particularly between de novo mutations and hard selective sweeps (Fig. 3A and SI Appendix, Table S6). Loci categorized as being under hard or soft selection had distinctive haplotype network structures, providing additional verification of our classification methods (Fig. 3B). For the 39 sweeps containing candidate cognition loci, 31% were classified as hard selective sweeps

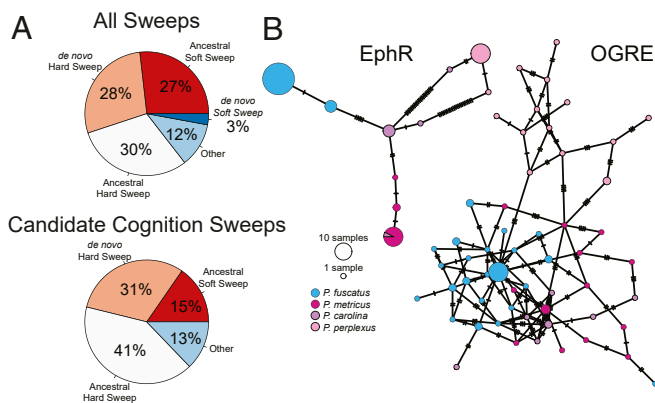


Fig. 3. (A) Selective sweep regions in *P. fuscatus* were classified as likely containing de novo alleles or ancestral alleles based upon the allele frequency distributions of *P. fuscatus* relative to the allele frequency distributions in four additional *Polistes* species. Candidate cognition selective sweeps were then classified as likely hard or likely soft sweeps using a machine learning approach. Both methods provide evidence for multiple hard sweeps of de novo mutations in *P. fuscatus*. (B) A representative haplotype network for a de novo/hard sweep (EphR) and for an ancestral/soft sweep (OGRE).

from de novo mutations, suggesting a contribution of hard selective sweeps during the evolution of cognitive abilities in *P. fuscatus*.

The Strongest Signatures of Selection on Cognition in *P. fuscatus* Are Predominantly Noncoding

We used two complimentary approaches to examine the relative importance of coding versus noncoding mutations on patterns of selection in *P. fuscatus*. First, we used snpEff (59) to classify the fixed and segregating SNPs detected in *P. fuscatus* as non-synonymous, synonymous, or noncoding mutations. Within the most extreme sweeps, we identified 476 SNPs with an allele frequency shift in *P. fuscatus* of 0.8 or greater, indicative of possible targets of selection. Of these, only 7 SNPs (1.5%) are nonsynonymous mutations. The nonsynonymous SNPs are found in 6 genes, 1 of which has a GO annotation suggesting a possible role in cognition. These data indicate that the strongest sweep signatures are likely to be associated with changes to noncoding elements. To directly compare these results with selection on protein sequences, we conducted a Bayesian implementation of the McDonald-Kreitman (MK) test using SNIPRE (60), comparing *P. fuscatus* to *P. metricus*. This method generated a per gene estimate of the average selection coefficient on nonsynonymous mutations (γ). We identified 52 genes with signatures of positive selection on their coding sequences ($\gamma > 1$). Genes were significantly enriched ($P < 0.05$) for seven GO terms (SI Appendix, Table S7), including detection of light stimulus involved in visual perception (GO: 0050908). While modest, these results provide some indication that selection on coding sequences may have contributed to the evolution of visual cognition. To understand the relationship between the results of the MK test and signatures of selective sweeps, we plotted the maximum CLR value for each genic locus against γ for that gene (Fig. 4). As might be expected, we find a positive correlation between these two values (Pearson's correlation: $r = 0.30$, $n = 8826$, $P < 2.2 \times 10^{-16}$), and this correlation is stronger when considering only genes under positive selection ($\gamma > 1$) (Pearson's correlation: $r = 0.64$, $n = 50$, $P < 3.4 \times 10^{-7}$). Notably, many loci with elevated CLR values show no evidence of selection on coding sequences. Taken together, these results argue strongly for the importance of regulatory evolution in *P. fuscatus* in general, as well as for the evolution of cognitive abilities in particular.

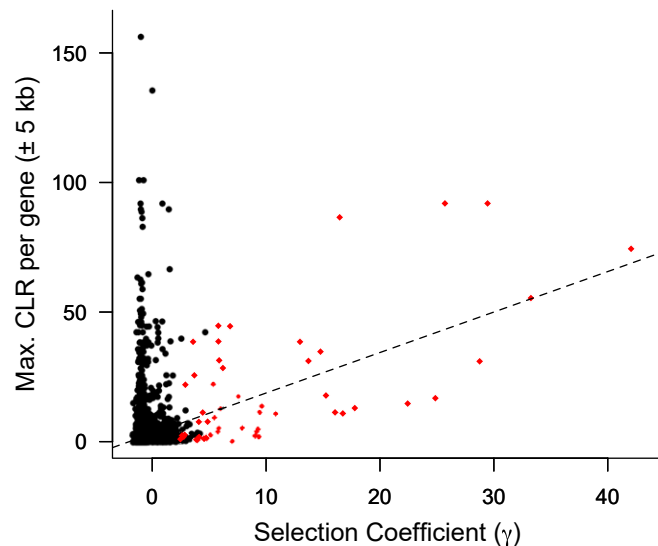


Fig. 4. Genes under positive selection ($\gamma > 1$) in *P. fuscatus* are positively correlated with regions of elevated composite likelihood ratio (CLR). Each point shows the value for a single *P. fuscatus* gene. Genes with $\gamma > 1$ are indicated in red. The selection coefficient (γ) was determined using a Bayesian implementation of the McDonald-Kreitman test. The maximum CLR per gene is the largest CLR value in a 1,000-bp window within the region spanning $\pm 5,000$ bp upstream and downstream of the gene. Many genes with high CLR values show no evidence of positive selection on the coding regions of the gene ($\gamma < 1$), suggesting recent selective sweeps on regulatory regions located near these genes.

Timing of Selective Events

To determine the relative timing of the detected selective sweeps, we generated estimates of coalescent times for 14 of the top candidate cognition loci with clear dominant haplotypes (Fig. 5). Estimated coalescent times for all selected alleles were recent, largely nonoverlapping, and occurred since the last glacial maximum. Since most of the alleles examined are implicated in hard selective sweeps from de novo mutations (Dataset S1), we infer that selection on each allele occurred at or very near the time the mutation emerged. Based on these distinct coalescent time estimates, we can reject a model by which recent selection on cognition in *P. fuscatus* took place predominantly in one time period. Instead, the data are consistent with a mutation-limited process in which cognitive evolution was constrained by the availability

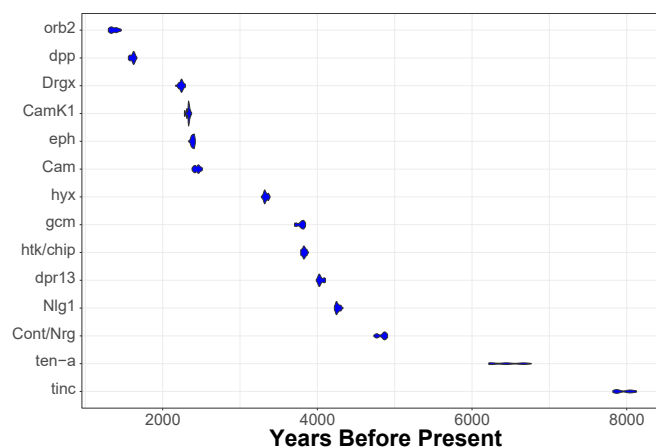


Fig. 5. Estimates of allele ages for several candidate cognition loci show evidence of several bouts of recent selection in *P. fuscatus*. Violin plots show estimates of the posterior distribution of the age of the most recent common ancestor of the allele.

of adaptive mutations, or a mutation order process where selective sweeps on beneficial mutations are contingent on the prior fixation of other adaptive alleles (61). We cannot rule out older selective events although multiple strong selective events since the last glacial maximum suggest a potential role for climate-induced shifts in range distribution, and cooperative behavior may have played a role in recent cognitive evolution in *P. fuscatus*, which extends further north than other North American *Polistes* wasps (27).

Mode and Tempo of Cognitive Evolution

Collectively, our findings are consistent with an oligogenic model (Fig. 1A) by which cognitive abilities in *P. fuscatus* evolved recently via strong selection on multiple loci of moderate to large selective effect. These findings should not be taken as evidence against weak selection on additional loci related to cognition, but rather as evidence for particularly strong signatures of selection on some loci during recent cognitive adaptation in *P. fuscatus*. These conclusions are robust to common criticisms of studies of selection in nonmodel organisms. First, it is often difficult to link selective sweeps to particular genes or regulatory regions, but, due to the high recombination rates in social Hymenoptera, we were able to detect narrow selective sweeps, often containing or nearby a single gene (Fig. 2A and Dataset S1). Furthermore, selective sweeps coincide with narrow peaks of divergence to a reconstructed ancestral genome, bolstering evidence for selection at these loci. Second, GO term analyses are generally used post hoc to explore large sets of genes, but here we used GO terms to test specific a priori hypotheses regarding selection on loci annotated with cognition-related GO terms in *P. fuscatus*. Third, linking candidate loci to specific functions or selective pressures is challenging, even for model organisms. Here, we used phylogenetic comparisons to show that the patterns of selection on loci related to learning, memory, visual processing, and cognition are unique to *P. fuscatus* (Fig. 2B), as would be expected if the selective signatures were related to the evolution of individual recognition. By comparing allele frequencies between *P. fuscatus* and four closely related species, we further demonstrate that, in many cases, selection appears to be acting upon *P. fuscatus*-specific de novo mutations (Fig. 3). We propose that a reasonable and parsimonious explanation for the patterns of selection on visual processing, learning, memory, and cognition detected in *P. fuscatus* is that the evolution of individual recognition in this lineage has been an important selective force in the species' recent evolutionary history.

The present study adds to a growing literature that identifies signatures of selection on nervous system and cognition annotated loci (20–23) but also provides insights on the potential for strong selection on de novo mutations during the process of cognitive evolution. We have identified three nonmutually exclusive hypotheses to explain why selection for cognitive abilities, potentially related to individual recognition, might generate such a strong signature of selection in *P. fuscatus*. First, the life history of paper wasps may favor strong selection for individual recognition. Cofoundress interactions in temperate zone paper wasp colonies have a large impact on fitness as successful cooperative nests have a higher likelihood of survival and greater productivity (62–64). Current evidence suggests it is these cofoundress interactions that favor individual recognition (25). Second, strong signatures of selection on cognition in our study may be related to timing. Plausibly, processes driving the initial evolution of a novel cognitive ability differ from the mechanisms maintaining cognitive variation within populations. Our data suggest that facial individual recognition in *P. fuscatus* may have evolved extremely recently (Fig. 5) whereas the initial evolution of important cognitive traits in many lineages occurred in the distant past, making it difficult to detect the dynamics of selection on loci related to those traits. Furthermore, our findings suggest, at least in part, that recent cognitive evolution in *P. fuscatus* may have been limited by the

availability or order of key adaptive mutations. Third, the dynamics we observed may be related to the fact that facial individual recognition is not only a novel cognitive trait in wasps but also a form of communication. In addition to cognitive changes, *P. fuscatus* also recently evolved increased color pattern diversity although these color pattern traits are expected to be under balancing selection and so would not be detected by the present analysis (65). If the evolution of novel communication systems explains the patterns observed here, then we might expect similar dynamics, including hard selective sweeps of novel mutations, to have played out for other novel communication systems as well, such as human language. Future studies of the tempo and mode of cognitive evolution in other species will help to clarify the generality of our findings.

Materials and Methods

Assembly and Annotation of Genomes. The *P. fuscatus* and *P. metricus* genomes were generated using PacBio SMRTbell libraries sequenced on the PacBio Sequel and 2 × 250-bp libraries sequenced on the Illumina HiSeq2500 sequencing platform. Reads were assembled with Canu (66), scaffolded with SSPACE-LongRead (67), and gap filled with Pilon (68). The *P. fuscatus* assembly was used as input for scaffolding by Dovetail Genomics. The *P. dorsalis* genome was constructed using the 10x genomics pipeline. The completeness of genome assemblies was assessed with BUSCO (69). The MAKER annotation pipeline (70) was used to identify predicted gene models, which were functionally annotated using the BLAST2GO workflow (71) and InterProScan (72). Detailed assembly and annotation methods are given in [SI Appendix](#).

Detection of Selective Sweeps. To identify genomic regions consistent with recent positive selection, we generated whole genome resequencing data from 40 *P. fuscatus*, 40 *P. metricus*, and 18 *P. dorsalis* individuals using 2 × 150-bp Illumina libraries. We identified SNPs using GATK (73) and determined allele frequencies with VCFtools (74). For each species, we used Sweepfinder2 (48) to infer selection based upon local deviations in the site frequency spectrum, generating a composite likelihood ratio test statistic score value (CLR) genome-wide in 1,000-bp windows. By using Discoal (75) to simulate a single selective sweep, we tested the effect of the strength of selection, timing of selection, and allele starting frequency on CLR value. We compared CLR to sequence divergence by using progressiveCactus (76) to reconstruct the genome of the ancestor of *P. fuscatus* and *P. metricus*. Program parameters are in [SI Appendix](#).

Selection on Cognition-Associated Gene Ontology Terms. Each gene was classified as a possible candidate locus for cognition if annotated with the following GO terms or offspring terms: cognition, mushroom body development, visual behavior, learning or memory, and eye development. All other genes were designated as unlikely a priori targets for cognitive evolution. We compared scaled CLR values among the three species for the two categories of genes. Additionally, for each of the three species, we used BLAST2GO (71) to look for enrichment of gene ontology (GO) terms in loci located within 5 kb of CLR outlier peaks and conducted a gene set enrichment analysis using the maximum CLR value for each genic locus.

Evolution of Coding and Regulatory Sequence. The predicted effect of each variant was determined using SnpEff (v4.3) (59). We performed a Bayesian implementation of the McDonald–Kreitman test in *P. fuscatus* using SnIPRE (60). Positions of synonymous and nonsynonymous mutations were determined from the SnpEff results. Parameters are given in [SI Appendix](#).

Classification of Selective Sweeps and Dating of Alleles. *P. fuscatus* SNPs were classified as likely de novo or ancestral by comparing the allele frequency from the 40 resequenced *P. fuscatus* genomes to the allele frequencies in four other closely related species (40 *P. metricus*, 18 *P. dorsalis*, 17 *P. perplexus*, and 18 *P. carolina*). We verified our classifications by using Discoal (75) and DiploSHIC (58) to detect and categorize regions under selection as hard or soft sweeps using a machine learning approach. For candidate cognition loci identified in hard selective sweeps, we estimated the age of the selected alleles using starTMRCA (42). Program parameters and simulations values are given in [SI Appendix](#).

ACKNOWLEDGMENTS. This work was supported by NIH Grant DP2-GM128202 (to M.J.S.), NSF CAREER Grant DEB-1750394 (to M.J.S.), and NSF Postdoctoral Fellowship DBI-1711703 (to S.E.M.). We thank K. Bessler for help with library preparation. T. Blankers, T. Hendry, A. Clark, A. Moeller, H. K. Reeve, K. Shaw, and A. Toth provided helpful comments on the project and manuscript.

1. S. J. Shettleworth, *Cognition, Evolution, and Behavior* (Oxford University Press, 2010).
2. S. D. Healy, S. R. de Kort, N. S. Clayton, The hippocampus, spatial memory and food hoarding: A puzzle revisited. *Trends Ecol. Evol.* **20**, 17–22 (2005).
3. S. M. Reader, K. N. Laland, Social intelligence, innovation, and enhanced brain size in primates. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4436–4441 (2002).
4. A. R. DeCasien, S. A. Williams, J. P. Higham, Primate brain size is predicted by diet but not sociality. *Nat. Ecol. Evol.* **1**, 0112 (2017).
5. F. J. Pérez-Barbería, S. Shultz, R. I. M. Dunbar, Evidence for coevolution of sociality and relative brain size in three orders of mammals. *Evolution* **61**, 2811–2821 (2007).
6. J. A. Basil, A. C. Kamil, R. P. Balda, K. V. Fite, Differences in hippocampal volume among food storing corvids. *Brain Behav. Evol.* **47**, 156–164 (1996).
7. T. S. Fristoe, A. N. Iwaniuk, C. A. Botero, Big brains stabilize populations and facilitate colonization of variable habitats in birds. *Nat. Ecol. Evol.* **1**, 1706–1715 (2017).
8. D. Sol, S. Bacher, S. M. Reader, L. Lefebvre, Brain size predicts the success of mammal species introduced into novel environments. *Am. Nat.* **172** (suppl. 1), S63–S71 (2008).
9. R. Croston, C. L. Branch, D. Y. Kozlovsky, R. Dukas, V. V. Pravosudov, Heritability and the evolution of cognitive traits. *Behav. Ecol.* **26**, 1447–1459 (2015).
10. B. J. Ashton, A. R. Ridley, E. K. Edwards, A. Thornton, Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature* **554**, 364–367 (2018).
11. A. S. Dunlap, D. W. Stephens, Experimental evolution of prepared learning. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11750–11755 (2014).
12. J. J. Lee *et al.*, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
13. B. Hare *et al.*, Social cognitive evolution in captive foxes is a correlated by-product of experimental domestication. *Curr. Biol.* **15**, 226–230 (2005).
14. B. Hare, M. Brown, C. Williamson, M. Tomasello, The domestication of social cognition in dogs. *Science* **298**, 1634–1636 (2002).
15. A. Kotschal *et al.*, Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Curr. Biol.* **23**, 168–171 (2013).
16. T. J. Polderman *et al.*, Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
17. M. Goodman *et al.*, Phylogenomic analyses reveal convergent patterns of adaptive evolution in elephant and human ancestries. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20824–20829 (2009).
18. M. R. McGowen, L. I. Grossman, D. E. Wildman, Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc. Biol. Sci.* **279**, 3643–3651 (2012).
19. M. Wirthlin *et al.*, Parrot genomes and the evolution of heightened longevity and cognition. *Curr. Biol.* **28**, 4001–4008.e7 (2018).
20. V. N. Laine *et al.*, Great Tit HapMap Consortium, Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7**, 10474 (2016).
21. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006). Erratum in: *PLoS Biol.* **5**, e147 (2007).
22. C. Chen *et al.*, Population genomics provide insights into the evolution and adaptation of the eastern honey bee (*Apis cerana*). *Mol. Biol. Evol.* **35**, 2260–2271 (2018).
23. X. Zhou *et al.*, Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat. Commun.* **9**, 1276 (2018).
24. M. Przeworski, The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).
25. E. A. Tibbetts, M. J. Sheehan, Individual recognition and the evolution of learning and memory in *Polistes* paper wasps. *Handb. Invertebr. Learn. Mem.* **22**, 561–571 (2013).
26. M. J. Sheehan, E. A. Tibbetts, Specialized face learning is associated with individual recognition in paper wasps. *Science* **334**, 1272–1275 (2011).
27. M. J. Sheehan *et al.*, Different axes of environmental variation explain the presence vs. extent of cooperative nest founding associations in *Polistes* paper wasps. *Ecol. Lett.* **18**, 1057–1067 (2015).
28. M. J. Sheehan, E. A. Tibbetts, Robust long-term social memories in a paper wasp. *Curr. Biol.* **18**, R851–R852 (2008).
29. H. K. Reeve, P. Nonacs, Social contracts in wasp societies. *Nature* **359**, 823–825 (1992).
30. J. K. Pritchard, J. K. Pickrell, G. Coop, The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
31. G. Bell, *The Oligogenic View of Adaptation in Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory Press, 2009), pp. 139–144.
32. M. Turelli, N. H. Barton, Dynamics of polygenic characters under selection. *Theor. Popul. Biol.* **38**, 1–57 (1990).
33. H. Bastide, J. D. Lange, J. B. Lack, A. Yassin, J. E. Pool, A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics* **204**, 1307–1319 (2016).
34. J. M. Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
35. J. Hermisson, P. S. Pennings, Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
36. R. D. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
37. N. L. Kaplan, R. R. Hudson, C. H. Langley, The “hitchhiking effect” revisited. *Genetics* **123**, 887–899 (1989).
38. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
39. H. E. Hoekstra, J. A. Coyne, The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
40. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
41. S. Nakagome *et al.*, Estimating the ages of selection signals from different epochs in human history. *Mol. Biol. Evol.* **33**, 657–669 (2016).
42. J. Smith, G. Coop, M. Stephens, J. Novembre, Estimating time to the common ancestor for a beneficial allele. *Mol. Biol. Evol.* **35**, 1003–1017 (2018).
43. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).
44. G. Bell, Evolutionary rescue and the limits of adaptation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120080 (2013).
45. S. Patalano *et al.*, Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13970–13975 (2015).
46. D. S. Standage *et al.*, Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Mol. Ecol.* **25**, 1769–1784 (2016).
47. L. Wilfert, J. Gadau, P. Schmid-Hempel, Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98**, 189–197 (2007).
48. M. DeGiorgio, C. D. Huber, M. J. Hubisz, I. Hellmann, R. Nielsen, SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897 (2016).
49. R. Gerlai, Eph receptors and neural plasticity. *Nat. Rev. Neurosci.* **2**, 205–209 (2001).
50. M. Vidovic, A. Nighorn, S. Koblar, R. Maleszka, Eph receptor and ephrin signaling in developing and adult brain of the honeybee (*Apis mellifera*). *Dev. Neurobiol.* **67**, 233–251 (2007).
51. A. Majumdar *et al.*, Critical role of amyloid-like oligomers of *Drosophila* Orb2 in the persistence of memory. *Cell* **148**, 515–529 (2012).
52. K. D. Curtin, Z. Zhang, R. J. Wyman, Gap junction proteins expressed during development are required for adult neural function in the *Drosophila* optic lamina. *J. Neurosci.* **22**, 7088–7096 (2002).
53. M. J. Sheehan, J. Jinn, E. A. Tibbetts, Coevolution of visual signals and eye morphology in *Polistes* paper wasps. *Biol. Lett.* **10**, 20140254 (2014).
54. W. Gronenberg, L. E. Ash, E. A. Tibbetts, Correlation between facial pattern recognition and brain composition in paper wasps. *Brain Behav. Evol.* **71**, 1–14 (2008).
55. M. J. Sheehan, E. A. Tibbetts, Selection for individual recognition and the evolution of polymorphic identity signals in *Polistes* paper wasps. *J. Evol. Biol.* **23**, 570–577 (2010).
56. J. Hermisson, P. S. Pennings, Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **8**, 700–716 (2017).
57. A. Avalos *et al.*, A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nat. Commun.* **8**, 1550 (2017).
58. A. D. Kern, D. R. Schrider, diploS/HIC: An updated approach to classifying selective sweeps. *G3 (Bethesda)* **8**, 1959–1970 (2018).
59. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
60. K. E. Eilertson, J. G. Booth, C. D. Bustamante, SnPRE: Selection inference using a Poisson random effects model. *PLoS Comput. Biol.* **8**, e1002806 (2012).
61. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
62. E. A. Tibbetts, H. K. Reeve, Benefits of foundress associations in the paper wasp *Polistes dominulus*: Increased productivity and survival, but no assurance of fitness returns. *Behav. Ecol.* **14**, 510–514 (2003).
63. D. L. Gibo, The selective advantage of foundress associations in *Polistes fuscatus* (Hymenoptera: Vespidae): A field study of the effects of predation on productivity. *Can. Entomol.* **110**, 519–540 (1978).
64. R. A. Metcalf, G. S. Whitt, Relative inclusive fitness in the social wasp *Polistes metricus*. *Behav. Ecol. Sociobiol.* **2**, 353–360 (1977).
65. M. J. Sheehan, J. Choo, E. A. Tibbetts, Heritable variation in colour patterns mediating individual recognition. *R. Soc. Open Sci.* **4**, 161008 (2017).
66. S. Koren *et al.*, Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
67. M. Boetzer, W. Pirovano, SSPAC-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
68. B. J. Walker *et al.*, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
69. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
70. C. Holt, M. Yandell, MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
71. A. Conesa *et al.*, Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
72. P. Jones *et al.*, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
73. A. McKenna *et al.*, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
74. P. Danecek *et al.*, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
75. A. D. Kern, D. R. Schrider, Discoal: Flexible coalescent simulations with selection. *Bioinformatics* **32**, 3839–3841 (2016).
76. B. Paten *et al.*, Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011).