Evolution of hunt, kill, devour in a *Vibrio* ecospecies

Sarah L. Svensson[1]*†, Huiqi Wen[1]†, Zhizhou Jia[1]†, Zhe Xie[1], Wenxuan Xu[1], Hui Fu[1], Chao Yang[1], Yanjie Chao[1]*, Daniel Falush[1]*

[1]Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, Shanghai, China 200031
†Equally contributing authors.
*To whom correspondence should be addressed. Email: sarah.svensson@ips.ac.cn, yjchao@ips.ac.cn, daniel.falush@ips.ac.cn

Abstract

Bacterial ecospecies are a recently recognized genetic structure that is hypothesized to arise by complex adaptation to a new niche. Using coadaptation analysis as a foundation for laboratory experiments, we find that the ecological strategy of the "Molassodon" ecospecies of *Vibrio parahaemolyticus* is to hunt, kill and devour bacterial prey in viscous liquids. Instead of swarming on surfaces, Molassodon strains have differentiated and repurposed their lateral flagella to enhance swimming. We functionally link Type 6 secretion and nutrient uptake to this motility phenotype and reconstruct the evolution of Molassodon, showing that several key genes have been imported from other *Vibrio* species. Our analysis confirms the functional coherence of bacterial ecospecies and provides a window into adaptation for environmental microbes, where characterizing relevant microhabitats and evolutionary strategies remains a considerable challenge.

Complex adaptations such as the flight of birds, the bacterial flagellum or the feeding strategies of orcas are amongst the most reliable sources of human wonder. However, establishing the genetic basis of these adaptations and the steps involved in their evolution is typically challenging due to the large number of genes and long time periods involved. In most cases intermediate genotypes are not available, limiting our ability to investigate the genotype to phenotype map in any detail.

To gain general insights into the process of complex adaptation, bacterial ecospecies are a promising new subject for investigation. A bacterial population with two ecospecies has a single-well mixed gene pool in most of the genome, but nearly-fixed differences between ecospecies at a small number of genes (*1*). We have proposed that the differences between ecospecies are maintained by natural selection, which implies that the genetic basis of adaptation can be established by identifying these fixed differences. The principal tasks instead become to characterize the phenotypic differences between ecospecies, the associated ecological strategies and the steps by which the differentiation evolved.

For example, *Helicobacter pylori* colonizes the stomachs of humans and occasionally other mammals. The Hardy ecospecies is differentiated from typical Ubiquitous *H. pylori* at only 100 of approximately 1500 genes and a handful of accessory genes (*1*). Based on the diets of humans and animals from which Hardy is isolated from as well as functional annotation of the differentiated genes, we have proposed that Hardy is a complex adaptation to host carnivory, a hypothesis that requires further functional validation. Analysis of the pattern of differentiation is consistent with Hardy evolving

progressively within the stomachs of ancestors of modern humans and then dispersing around the world with us.

Some strains of *Vibrio parahaemolyticus* can also briefly colonize humans, leading to gastroenteritis (*2*). However, unlike *H. pylori, V. parahaemolyticus* is mostly free-living in oceans worldwide but is also associated with diverse marine animals, together with other *Vibrio* species. This broader environmental niche allows a wider source for its gene pool than *H. pylori*, which rarely comes into contact with other *Helicobacter* species. Several of these *Vibrio* species have served as models for studying bacterial motility and also ecologically related functions such as Type 6 secretion-mediated bacterial killing (*3*–*6*). We previously identified a highly distinct group of *V. parahaemolyticus*, EG1a, characterized by differentiation at genes including lateral flagella (*6*), secondary flagella usually used for swarming motility on surfaces by all *V. parahaemolyticus* and a small subset of other *Vibrio* spp. (*7*). While the genotype is globally distributed and represents a significant percentage of strains in the database, its ecological niche and how differentiation might be associated with this was unclear.

Here, we show that EG1a is a bacterial ecospecies. We identify and refine a distinct laboratory phenotype, finding that unlike other *V. parahaemolyticus*, they turn on their lateral flagella when swimming through viscous liquids, achieving faster dispersal, but that they have lost the ability to swarm. Accordingly, we rename the ecospecies "Molassodon", with strains lacking the genotype described as Typical. Additional Molassodon genes are involved in bacterial killing and nutrient acquisition, leading us to infer that the ecological niche of the ecospecies is foraging bacterial prey in a viscous environment. We also reconstruct steps by which this phenotype evolved, establishing a key role for introgression of genes from other *Vibrio* species. Our study opens new avenues to explore the genetic basis of complex adaptation and the ecological factors driving it.

Results

The Molassodon ecospecies evolved by introgression

We previously identified a multilocus genotype in *Vibrio parahaemolyticus* (Molassodon - previously EG1a), which we here show constitutes an ecospecies. Based on only strains from VppAsia, we had previously used a GWES (genome-wide epistasis scan) to identify Molassodon based on hierarchical clustering of a large group of interacting genes (*8*). We here expand on this with a GWAS (genome-wide association study) of Molassodon vs. Typical strains using a larger set of non-redundant genomes representing all four geographical populations (*9*) to identify SNPs differentiating Molassodon from Typical (**Fig. 1A; Tables S1-S4; Fig. S1A**). Pangenome GWAS also identified a large set of accessory genes significantly associated with Molassodon/Typical, distinguishing the ecospecies structure from that of *H. pylori* Hardy (**Fig. 1B**; **Table S4**). These analyses confirmed that Molassodon is an ecospecies, differentiated by only 48/~4800 core genes and over 70 accessory genes.

Trees generated from differentiated core genes and a set of undifferentiated genes (**Table S3**) showed different topologies, as we saw previously for *H. pylori* Hardy (*1*). The undifferentiated gene tree was

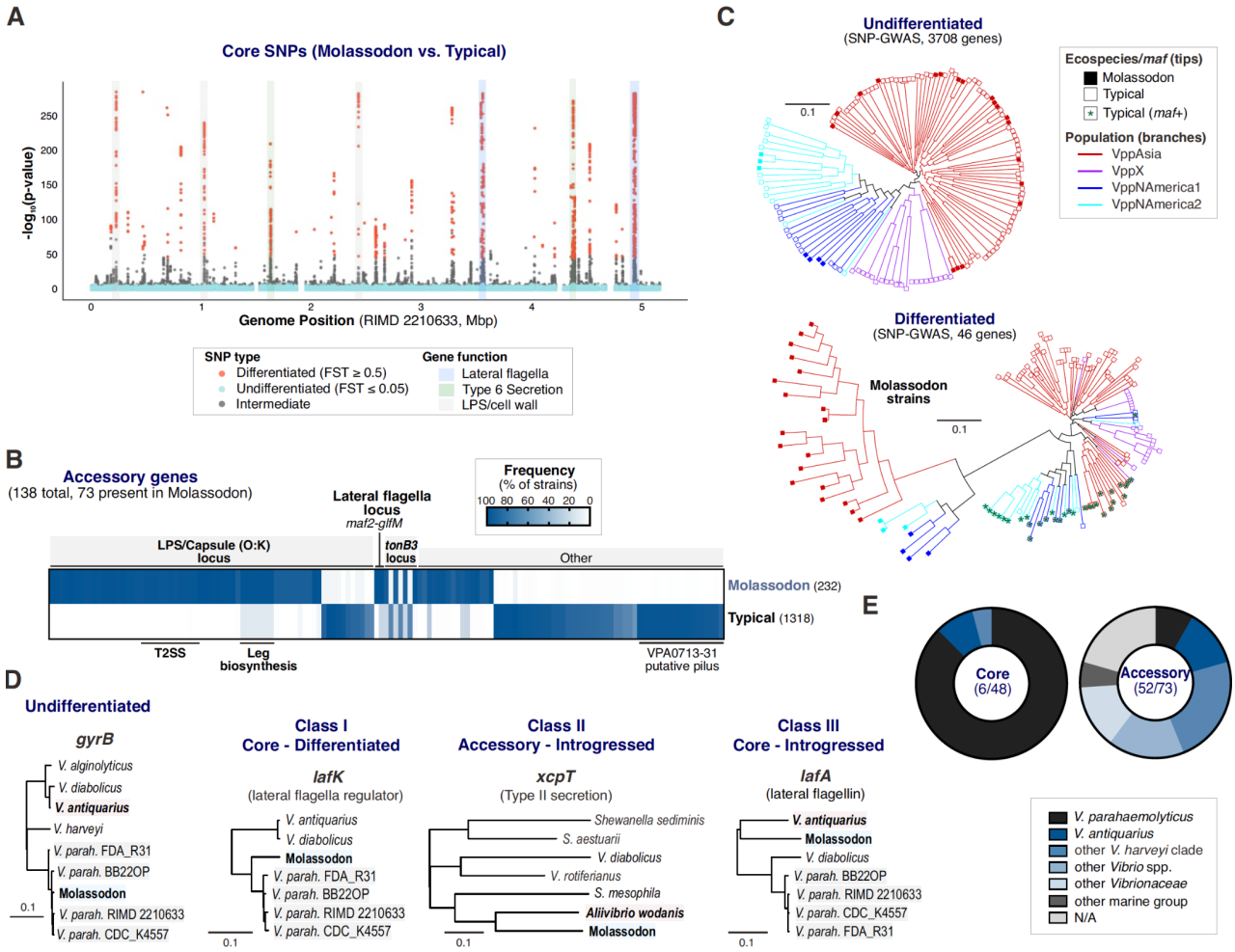

Figure 1. The Molassodon ecospecies genotype. (A) Manhattan plot for core genome SNP (single-nucleotide polymorphism) GWAS of Molassodon (n=223) vs. Typical (n=1318) strains. Significant SNPs: FST (fixation index) ≥0.5 and -log10(p) ≥10. See also **Table S2. (B)** Frequency of genes identified as significantly differentially distributed between Molassodon and Typical strains via pangenome GWAS. See also **Table S4**. Significant genes: FST ≥0.5 and -log10(p) ≥ 10, also requiring >90% or <10% frequency in Molassodon. Genes are ordered by putative function. **(C)** Neighbour-joining trees for Undifferentiated (*top*, 3708 genes) and Differentiated genes (*bottom,* 46 genes) identified in core genome SNP GWAS. A subset of 137 strains with representatives of each group of interest (population, ecospecies) was used. Differentiated genes: at least five differentiated SNPs. Undifferentiated genes: no SNPs with FST ≥0.05. Branches: geographic population. Tips: ecospecies (Typical/Molassodon) and *maf* (*maf2-glfM* accessory gene) genotype. See also **Table S3** for Differentiated and Undifferentiated genes. **(D)** Representative trees for KEGG matches for undifferentiated (*gyrB*) and representative examples of three classes of Molassodon-differentiated genes. Class I: core-genome genes with closest match in other *V. parahaemolyticus* strains, but with an extended branch length. Class II: accessory genes with signs of introgression (*i.e.*, closest match outside of *V. parahaemolyticus*). Class III: core-genome genes with signs of introgression. Grey: Typical *V. parahaemolyticus*. Blue: Molassodon. Pink: Best species match in the database. For additional examples, see **Fig. S3**. **(E)** Taxon of best KEGG ortholog (by % amino acid identity) for Molassodon-present accessory and core genes. N/A: no match above 40%.

star-like and reflected geographic populations (*9*), with Molassodon strains from three populations dispersed amongst Typical strains (**Fig. 1C,** *top*). Based on differentiated genes, all Molassodon strains instead clustered together on an extended branch, no matter which population they were from (**Fig. 1C,** *bottom*). Therefore, Molassodon genes have a different evolutionary path, despite extensive ongoing genetic exchange between ecospecies in most of the genome.

Further inspection showed that there are three classes of Molassodon-differentiated genes based on their mode of evolution (**Figs. 1D & 1E; Fig. S2**). Most core genes are evolving by point mutation with a clear signature of branch-lengthening vs. other *V. parahaemolyticus* (e.g., *lafK*; **Fig. 1D**), implying accelerated evolution of the Molassodon version of the gene. In contrast, most accessory genes (52/73) show hallmarks of introgression from other species (e.g., *xcpT*; **Fig. 1D**). However, we also identified six core genes with evidence of allelic exchange with other species, including *lafA*, encoding the lateral flagellin, and several genes in the *tonB3* transporter locus (**Fig. 1D & Fig. S2B)**. The most common closest match for introgressed Molassodon genes was *V. antiquarius*.

As we previously noted (*8*), many of the imported Molassodon accessory genes are encoded in a large genomic block, and here we report that this block is the single locus encoding the LPS/capsule (O:K) genes in *Vibrio* spp. This locus is the basis of O:K serotyping in the species (*10*) and despite high variability in our dataset, we found that the O:K locus of Molassodon is present in only a handful of versions (**Fig. S3A & S3B; Table S5**). Most fixed O:K genes also had hallmarks of introgression (**Table S4**).

In sum, this implies that a large proportion of the multigene Molassodon genotype has been acquired from outside of *V. parahaemolyticus*, making the ecospecies distinct from the Hardy ecospecies of *H. pylori,* which has primarily evolved via sequence divergence.

Molassodon strains do not swarm despite inducing lateral flagella

Because of the coinheritance of Molassodon genes in an otherwise well-mixed species, indicating they are under selection, we presume that Molassodon encodes a complex adaptation. However, this raises the question of what the adaptation is. Our goal here is to identify a laboratory phenotype that is as informative as possible about what the adaptation is and how it works in the natural environment. Our biggest clue to a tractable phenotypic difference between Molassodon and Typical strains was motility, based on differentiation of lateral flagella genes (**Fig. 1A**). In fact, we found that Molassodon strains do not swarm on surfaces like typical *V. parahaemolyticus*.

Lateral flagella are a secondary set of environmentally-regulated, numerous, peritrichous flagella that mediate *V. parahaemolyticus* swarming motility on surfaces (*11–13*). This is in addition to polar flagella, which are constitutively expressed and mediate swimming in liquids. Not only structural components (flagellin, hook, and cap proteins; several components of the basal body and motor) but also several lateral flagella regulators are very different in Molassodon strains (**Fig. 2A**). For example, the most differentiated core lateral flagella gene encodes a key regulator involved in their induction, LafK. This foreshadowed not only a functional difference, but a potential difference in environmental regulation of lateral flagella.

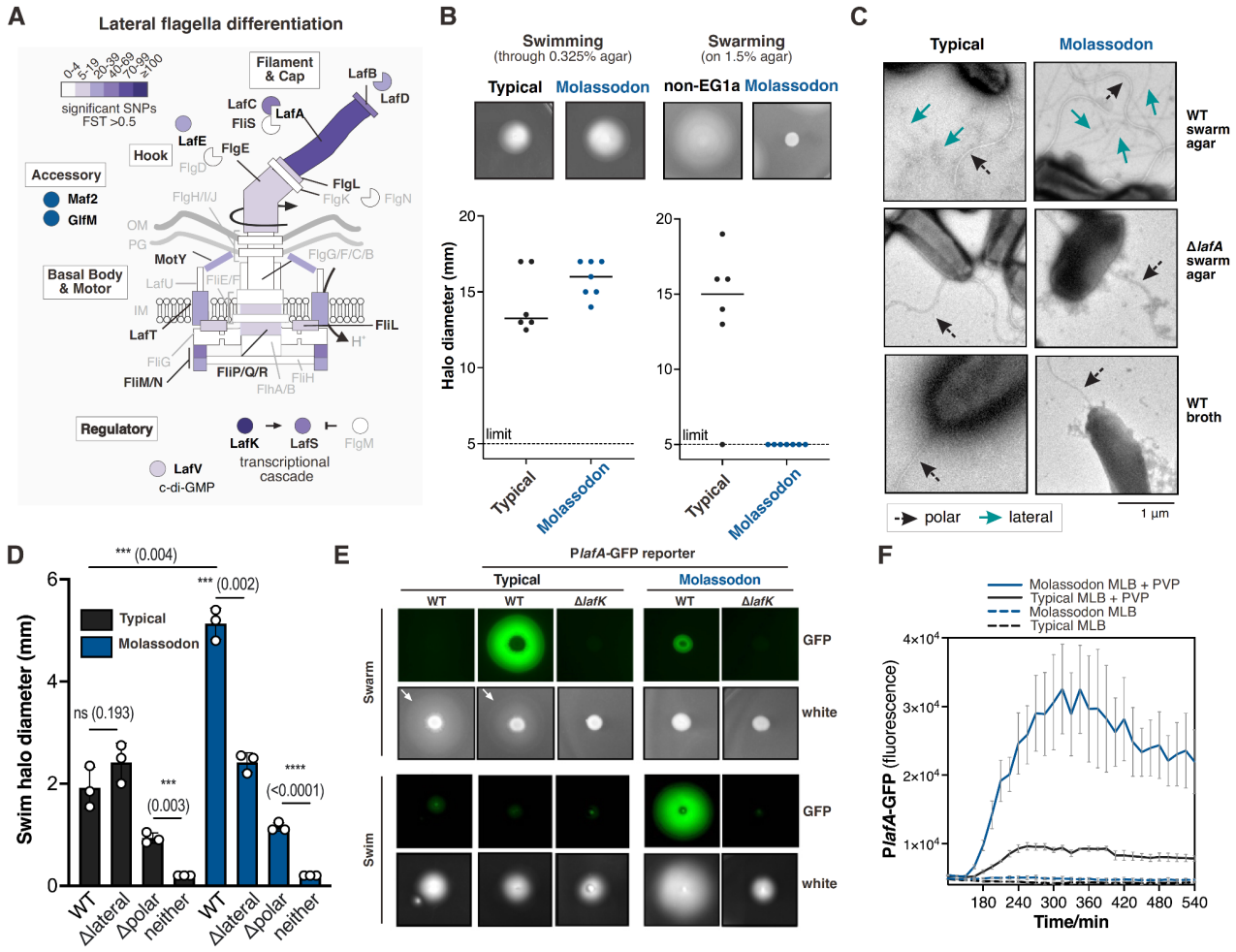

Figure 2. Molassodon uses lateral flagella for swimming in viscous media rather than swarming. (A) Schematic of lateral flagella differentiation in Molassodon strains. Differentiated: at least five differentiated SNPs (FST ≥0.5 and -log10(p) ≥10); Molassodon vs Typical GWAS. **(B)** *Top*: Representative swim/swarm halos for Typical and Molassodon strains. *Bottom*: Halo measurements. Representative of at least two independent experiments. **(C)** Transmission electron micrographs of Typical and Molassodon strains on swarm agar. Δ*lafA*: deletion of lateral flagellin gene (VPA1548). Typical WT was harvested from the outside of swarm halos; Molassodon WT and Δ*lafA* for both ecospecies were recovered from the macrocolony that forms in the absence of swarming (*e.g.*, Molassodon, panel B). Arrows: polar (black) or lateral (green) flagella. **(D)** Swim assay in low concentration (0.25%) agar for Typical and Molassodon WT or polar/lateral flagella deletion strains. Δlateral: Δ*lafA*, lateral flagellin. Δpolar: Δ*fliS*, polar flagellin chaperone. Neither: Δ*lafA* Δ*fliS*. *** $p < 0.001$, **** $p < 0.0001$, ns - $p > 0.05$, Student's *t*-test. **(E)** Activity of *lafA* promoter reporters (~200 nt upstream of the transcriptional start site) in Molassodon/Typical backgrounds on swarm agar (*top*) and in swim agar (*bottom*). Representative of at least three independent experiments. **(F)** Expression of P*lafA* reporters in Molassodon and Typical in the absence (MLB) and presence (MLB+3% PVP, polyvinylpyrrolidone). Error bars: standard deviation for three strains of the same ecospecies. See also **Fig. S4D**.

We compared motility (swimming and swarming) for a panel of natural Molassodon vs. Typical (which includes most typical *V. parahaemolyticus*, including the reference RIMD2210633) isolates. Molassodon strains produced moderately larger swim halos than Typical strains when inoculated

into standard swim agar plates (**Fig. 2B**, *left*). In contrast, none of the Molassodon strains showed any migration when spotted on the surface of swarm plates, and instead grew as a macrocolony (**Fig. 2B**, *right*). We tested a variety of temperatures, agar concentrations, agar sources, salt concentrations, surfactants, and genetic changes (quorum sensing, c-di-GMP, transposon suppressor screen), but puzzlingly did not observe any migration of Molassodon on a surface (data not shown). Despite absence of swarming behaviour, RNA-seq showed that Molassodon strains still induced lateral flagella genes on a surface, although the level of induction was generally 10-fold lower than for Typical strains (**Fig. S4A; Table S7**). Nonetheless, Molassodon strains can build visible lateral flagella when grown on a surface (green arrows, **Fig. 2C**), albeit without an obvious function.

Molassodon uses lateral flagella for swimming in complex environments

While trying to understand this apparently "negative" phenotype, we found that Molassodon strains uniquely induced lateral flagella during swimming in low viscosity swim media. Our next experiments sought to show that Molassodon lateral flagella retained a motility function, by observing motility in the absence of polar flagella. Accordingly, we tested migration under "less challenging" (very low-viscosity swim media, 0.25% agar) conditions. Both Molassodon and Typical strains expressing only lateral flagella (Δ*fliS*, VP2254) produced small swim halos, which disappeared upon further deletion of *lafA* (**Fig. 2D**, bars 3,4 and 7,8). We thus concluded that Molassodon lateral flagella are in fact functional for motility.

The control strains in this experiment revealed a positive Molassodon motility phenotype. Above, we observed that Molassodon strains generally swam better than Typical strains (**Fig. 2B,** *left*), and in lower percentage agar (0.25% vs. 0.325%), this was much more pronounced and significant (**Fig. 2D**; 1st vs 5th bars). Strikingly, the larger swim halos of WT Molassodon were reduced 2-fold by *lafA* deletion (**Fig. 2D**; 5th vs 6th bars). In contrast, Δ*lafA* did not affect Typical swimming. This is expected: *V. parahaemolyticus* lateral flagella are not thought to be substantially expressed in swim assays. However, the mutant analysis suggested otherwise for Molassodon.

To confirm that the Molassodon strain was inducing lateral flagella during swimming, we generated a *lafA* promoter-GFP reporter for each ecospecies (due to minor differences in promoter sequence) (**Fig. S4B**). As expected, Typical cells induced P*lafA* only on swarm agar (**Fig. 2E**). In contrast, P*lafA* was active in swimming Molassodon. We applied RNA-seq to three Molassodon and three Typical strains harvested from swim agar and confirmed that *lafA* promoter induction reflected higher expression of all lateral flagella transcripts in the Molassodon strains (**Fig. S4C; Table S8**). Together with the effect of Δ*lafA* on Molassodon swimming (**Fig. 2D**), this experiment suggests that Molassodon was using its lateral flagella to swim through complex media. Moreover, as viscosity is a known signal for induction of *V. parahaemolyticus* lateral flagella (*14*), we tested the effect of adding PVP (polyvinylpyrrolidone), previously used to explore the effect of viscosity on motility in several *Vibrio* species (*3*, *14*, *15*), on induction of P*lafA*. All three Molassodon strains induced P*lafA* more strongly than the three Typical strains in response to PVP (**Fig. 2F; Fig. S4D**). Overall, these experiments reveal the first laboratory phenotype for a bacterial ecospecies and suggest that Molassodon uniquely induces and uses its lateral flagella for swimming in viscous environments.

Hunt, kill, devour phenotype of Molassodon

The identification of a "positive" Molassodon laboratory phenotype, namely larger swim halos in low-density agar, raised the question of how closely related this phenotype is to the complex adaptation of Molassodon strains. More specifically, is a key Molassodon trait enhanced swimming and if so, what are the circumstances in which it brings an advantage? Our essentially complete list of the differentiated genes involved in the complex adaptation allows us to probe these questions using the tools of molecular microbiology. Tn-seq allows us to probe which of the Molassodon-differentiated genes contribute to the enhanced swimming phenotype, while RNA-seq measures which genes are co-expressed in this condition (**Fig. 3A**), potentially providing additional clues about the associated ecological context and the molecular details of the complex adaptation.

We used Tn-seq to measure Molassodon and Typical transposon insertion mutant pools either isolated from the outside of swim halos or grown in the same medium with shaking to remove a requirement for motility (**Fig. 3A**). As an additional control, we also measured the Typical strain from the outside and inside of swarm halos (**Fig. S5A**). Tn-seq information was integrated with RNA-seq differential expression (swim/swarm vs. LB) for each gene. In total, 76/116 Molassodon genes (66%) that we could assay using both methods were involved in the Molassodon swimming phenotype (**Fig. 3A**), with 39/116 genes induced, 56 required, and a 19 gene overlap (**Table S9**; **Fig. S5B & S5C**). Molassodon genes were enriched in both swim-induced and swim-required gene sets (padj=5.36e-16, normalized enrichment score = 3.36 / padj=3.82e-12, normalized enrichment score = -2.68, respectively). Therefore, the phenotype encompasses most Molassodon-differentiated genes. Further, lateral flagella genes were required, as expected for Typical swarming (**Fig. S5C; Table S11**) but not swimming (**Fig. 3A; Table S10**). These results are consistent with lateral flagella acquiring a new function in Molassodon strains.

To extend the Molassodon phenotype, we next focused on Molassodon genes that do not encode lateral flagella. Two functional classes of genes were co-induced with lateral flagella during Molassodon swimming, although they did not have a Tn-seq phenotype: the T6SS2 Type 6 secretion system, and the TonB3 transport system (**Fig. 3A & Fig. S6A**). Both were also induced to some extent in Typical strains during swimming and swarming (**Fig. S5B**). T6SS are contact-dependent killing systems that can be used for interbacterial competition, and several structural components and proteins that control assembly/disassembly of T6SS2, encoded by almost all *V. parahaemolyticus* strains, have differentiated in Molassodon (**Fig. 3B**). The most differentiated T6SS2 gene encodes the effector RhsP (*16*). Because motility can increase T6SS efficacy (*17–19*), we hypothesized that this might underlie co-adaptation and co-expression of these functions in Molassodon. We took advantage of specific induction of lateral flagella by PVP in Molassodon (**Fig. 2F**) to determine if lateral flagella-driven motility might impact T6SS2-mediated killing. We found that only in Molassodon did *lafA* deletion impact killing of prey *E. coli* (**Fig. 3C**). These results, together with co-evolution of T6SS2 and lateral flagella genes, imply that the Molassodon lateral flagella-dependent swimming phenotype facilitates killing of bacterial prey.

The second co-expressed function, TonB, provided a clue to the functional advantage of linking motility and interbacterial competition: hunting. *V. parahaemolyticus* strains, including Molassodon,

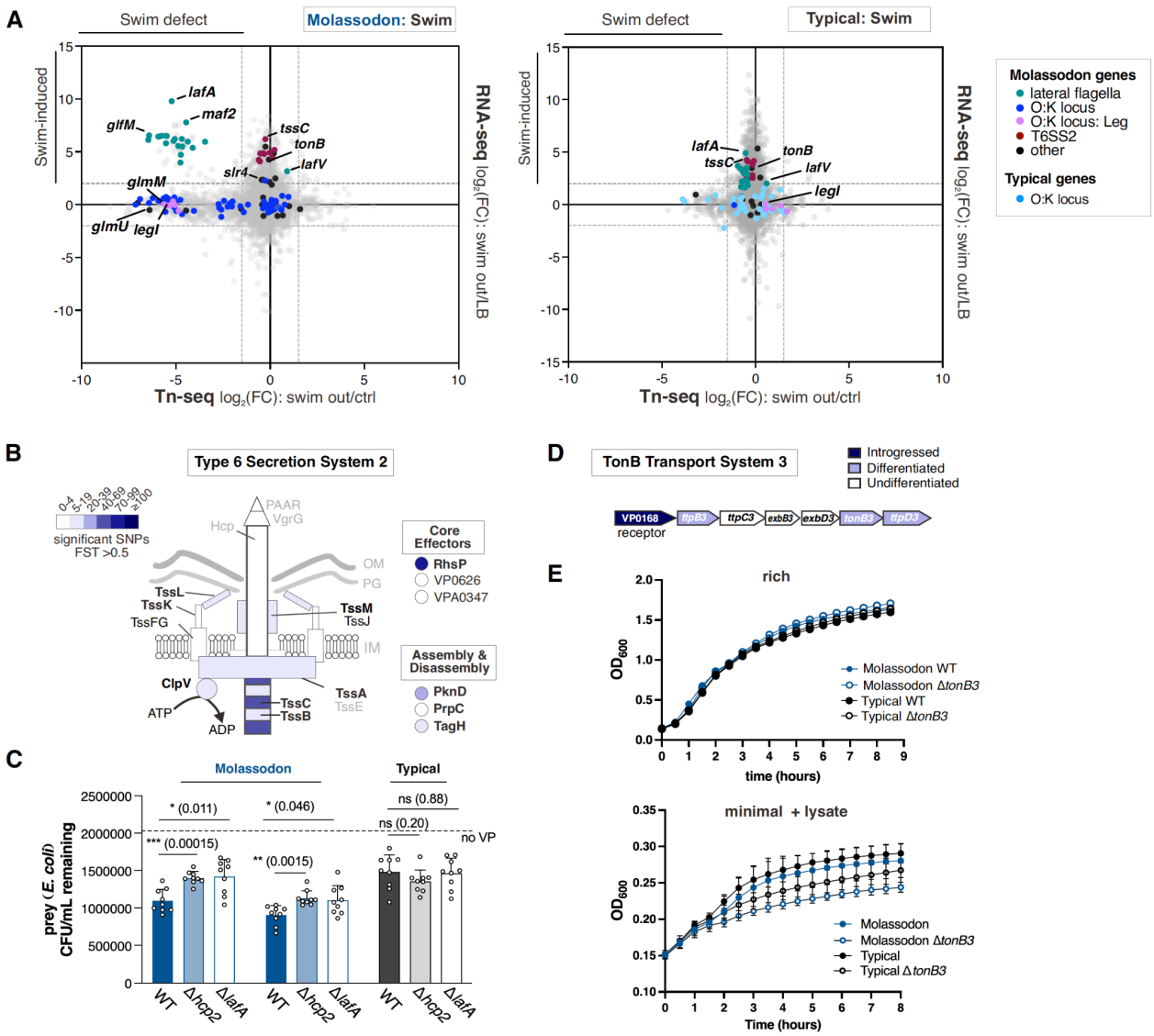

Figure 3. Validation and refinement of the Molassodon laboratory phenotype with Tn- and RNA-seq. (A) Tn-seq and RNA-seq profiles for Molassodon and Typical strains during swimming. Tn-seq: RIMD 2210633 (Typical) and DFS-0010 (Molassodon), $\log_2$FC Swim halo vs. Swim control. RNA-seq: mean of three Typical or Molassodon strains, $\log_2$FC Swim halo vs. LB. Coloured dots: Molassodon genes based on GWAS; Grey: all other genes shared by the ecospecies. See also **Fig. S5B** and **Tables S9-10**. **(B)** Schematic of T6SS2 with differentiated components indicated. **(C)** Killing activity against *E. coli* BW25113 for Typical and Molassodon wild-type, *lafA*, and T6SS2 (Δ*hcp2*) mutants in M9 broth (with 2% NaCl) + 5% PVP. The colony-forming units of the *E. coli* strain were determined after 3.5 hours of incubation at 30°C and normalized to the number for the corresponding *E. coli* group without added attacker for each ecospecies. Dashed line: mean prey CFU/ml remaining without added *V. parahaemolyticus* attacker *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ns not significant, Student's *t*-test. **(D)** Schematic of differentiated genes encoding a TonB/ExbBD transporter and receptor in Molassodon. **(E)** Molassodon strains grow better than Typical strains **(E)** Deletion of differentiated *tonB3* affects growth of a Molassodon, but not Typical, strain when prey (*E. coli*) lysate is provided as a nutrient source. n=3.

encode three TonB systems (TonB1-TonB3, according to *V. vulnificus* nomenclature (*20*)) (**Fig. 3D**), which transduce the energy of the proton-motive force for active transport of specific nutrients (iron compounds, vitamins, ions, sugars) across the outer membrane. Several components of TonB3, whose specificity is so far unknown, are differentiated in Molassodon, although there are no fixed differences with Typical strains (**Table S2**). There is evidence for introgression of several TonB3 components and the receptor is especially divergent (**Fig. S2B**), suggesting there could be specificity differences. To determine if TonB3 can facilitate nutrient acquisition from bacterial prey, we grew Molassodon and Typical WT/Δ*tonB3* (VP0163) strains in M9 minimal medium salts with only lysate from *E. coli* added as a carbon/nutrient source. We did not observe a clear trend for growth of WT strains (**Fig. S6**), but all Δ*tonB3* mutants reached lower densities than their respective parental wild-type strains (**Fig. 3E**). This confirms that TonB3 can mediate uptake of lysate components that can be used for growth by both ecospecies. Together, these experiments suggest a regulatory, and potentially also functional, difference between Molassodon and typical *V. parahaemolyticus* in three co-adapted pathways (lateral flagella-driven motility, Type 6 secretion, and nutrient uptake) that are co-induced under viscous conditions and together help *V. parahaemolyticus* to grow faster by using bacterial prey as a nutrient source.

Molecular and evolutionary assembly of Molassodon

In addition to lateral flagella, many other Molassodon genes were implicated by Tn-seq in the motility phenotype (**Fig. S7A**) and we next investigated why. For example, several regions of the Molassodon LPS/capsule island, which show hallmarks of introgression from outside of (Typical) *V. parahaemolyticus*, had significant Tn-seq phenotypes (**Fig. S7A**). Molecular dissection explained this functional linkage, as well as providing first insights into the stepwise assembly of the Molassodon genotype.

The Molassodon lateral flagella locus encodes two accessory genes, *maf2-glfM* (hereafter "*maf"),* which are absent in Typical strains and required for/induced during Molassodon swimming (**Fig. 3A & 4A**). Along with upstream *lafA*, the genes show hallmarks of introgression (**Fig. 1D & Fig. S7B**). Based on homology to *Shewanella* genes (*21*), as well as co-inheritance with *lafA* and predicted complex formation (**Fig. 4B**), Maf2 and GlfM likely act together to glycosylate Molassodon LafA with a sialic acid sugar. Consistent with this inference, a cluster of Molassodon genes in the OL2 LPS island encodes a pathway for biogenesis of the sugar legionaminic acid (Leg) (**Fig. 4C**), and O2 serotype LPS includes a Leg derivative (*22*). Importantly, each of these genes, as well as two Molassodon-differentiated core genes that feed into this pathway (*glmM*, *glmU*) were also required for the Molassodon fast swimming phenotype, but were dispensable for swimming in Typical (**Fig. 3A; Fig. S7A)**. To demonstrate that the Leg pathway genes work together with *maf2-glfM* and *lafA*, we generated double and triple deletion mutants. These mutants phenocopied Δ*lafA* for swimming, suggesting they affect motility in the same pathway (**Fig. 4D**). Therefore, our data explain the functional linkage between accessory Molassodon genes, and also provide an evolutionary timeline for assembly with biogenesis of legionaminic acid and LPS modification likely preceding flagellin glycosylation.

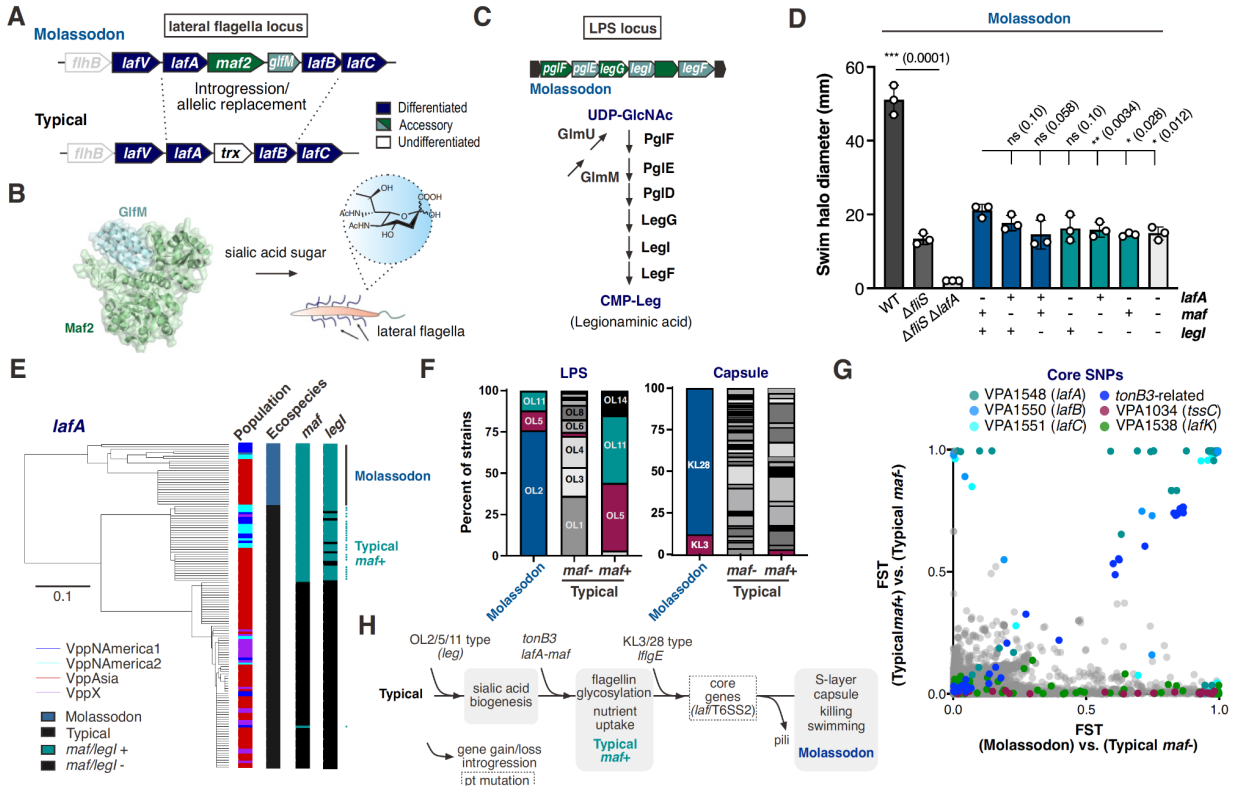

Figure 4. Convergent coadaptation after *maf* introgression. (A) Genomic context of the Molassodon accessory lateral flagella genes *maf2-glfM* and related region in non-introgressed strains. **(B)** Predicted complex formed by Maf2-GlfM and role of the proteins in flagellin glycosylation. The complex was predicted with Alphafold Multimer at Colabfold (*23*, *24*) and visualized with Pymol. **(C)** Molassodon genes encoding a putative Leg biosynthesis pathway. The pathway is based on (*25*)). **(D)** Motility assays with *maf2* and *legI* mutants. *Left*: Epistasis analysis of the contribution of Maf (*maf2-glfM*), Leg biosynthesis (*legI*), and lateral flagella (*lafA*) on swimming motility in Molassodon. *Right*: Effect of *legI* deletion on swimming in a Typical strain. *** p <0.001, ** p<0.01, * p<0.05, ns: not significant, Student's *t*-test, n=3. **(E)** Phylogenetic tree based on *lafA* for a subset of 137 strains with presence of "*maf*" (*maf2-glfM*) indicated. **(F)** Predicted serotypes for representative strains separated into Molassodon, Typical *maf*-, and intermediate (Typical *maf*+). Key O- or K-types are indicated with colour. See also **Table S5**. **(G)** Correlation between FST (fixation index) values for (Typical *maf*+) vs. (Typical *maf*-) and (Molassodon) vs. (Typical *maf*-) comparisons for the core genome SNPs. Key functions identified in this and previous (*8*, *26*) work are highlighted. For accessory genes, see **Fig. S10A**. **(H)** Evolutionary model for assembly of Molassodon.

How might such a multigene, complex adaptation have evolved? We noticed that *maf* is present in 15-25% of Typical strains in our dataset (**Fig. 1C**; **Table S4**). The *maf* genotype was also associated with presence of the Leg biosynthesis genes (**Fig. 4E**), consistent with a related function in glycosylating flagella. As the Leg genes are encoded in the LPS locus, we compared the predicted serotype composition of *maf*- vs. *maf*+ Typical strains. Molassodon strains encoded represent only three O-types (OL2, OL5, OL11), in comparison to Typical *maf*- strains, which showed high variability and diversity (**Fig. 4F**). All three O-types encode Leg biosynthesis genes (**Fig. S10B**). Typical *maf*+ strains also had limited O-locus diversity, with two types overlapping with Molassodon (OL5 and OL11) and a single type (OL14) shared with only one Typical *maf*- strain. However, the version of *maf* in Typical

strains, as well as *lafA*, has been independently derived (**Fig. 4E**), and is more like a version of the gene found in *V. antiquarius* than to the Molassodon version (**Fig. S8C**).

In addition to carrying Leg biosynthesis genes, most *maf*+ Typical strains are in an intermediate position between Molassodon and Typical strains in a tree based on Molassodon-differentiated core genes (**Fig. 1C**). We therefore performed a GWAS to identify which part of the Molassodon genotype they shared (**Fig. S8A & S8B; Table S2**). Shared fixed differences between Typical *maf*- and *maf*+ strains were restricted to one gene block surrounding *maf-glfM*, including a proportion of SNPs in *lafA*, as well as the downstream genes *lafB* and *lafC* (**Fig. 4F**). Shared non-fixed core genome differences between *maf*+ and *maf*- non Molassodon strains also included several in the operon encoding the TonB3 transporter investigated above (**Fig. 4F**). Interestingly, Molassodon and *maf*+ Typical strains both have a similar high frequency of a version of the *tonB3* operon, including the highly differentiated receptor, which is rare in *maf*- strains and is also partially introgressed (**Fig. 4F; Fig. S9A & S9B**) and was implicated in the devour phenotype above (**Fig. 3E**). In the accessory genome, we found that T6SS1 is absent in Molassodon, found at low frequency in Typical *maf*+ strains, and at high frequency in *maf*- strains. As previously observed (*26*), the cellulose biosynthesis cluster is in strong negative LD (linkage disequilibrium) with T6SS1, and accordingly shows an opposite pattern of frequencies (**Fig. S10A**).

Notwithstanding the similarities with Molassodon strains in *maf* and LPS, we found that all intermediate strains could swarm (**Fig. S10B**) and did not have obviously enhanced swimming. The positive phenotype of the *maf*+ Typical strains is therefore currently unclear and is presumably more subtle than that of Molassodon. However, we wondered if *de novo* introduction of *maf* into a hitherto *maf*- strain might produce a "hopeful monster" motility phenotype that was more obvious in laboratory conditions. Accordingly, we introduced *maf* on a plasmid into a *leg*+ Typical strain and found that it had significantly enhanced swarming (**Fig. S10C**), confirming that *maf* acquisition alone can affect lateral flagella-dependent motility.

While present-day *maf*+ intermediate strains may have recapitulated the initial stages of ecospecies evolution, they are indistinguishable from other Typical strains at most ecospecies loci (**Fig. 4G; Fig. S8A**). Molassodon has undergone more extensive divergence in lateral flagella loci, encompassing 233 SNPs in 15 additional genes to the three genes differentiated between *maf*- and *maf*+ strains. In particular, the LafK lateral flagella regulator has not yet differentiated in intermediate strains, suggesting, consistent with our phenotypic data (**Fig. S10B**), that distinct environmental regulation may not have yet evolved. The Molassodon specific lateral flagella differentiation at LafK and other genes appears to have been by point mutation, rather than further introgression.

Differentiation at the T6SS2 involved in our killing phenotype is also restricted to Molassodon strains. Although there is overlap between Typical and Molassodon in serotypes, the capsule locus shows a distinct pattern (**Fig. 4F**), consistent with a subsequent remodelling. Molassodon has a Kps-type capsule, expressed via an ABC transporter system (*27*), and perhaps most strikingly, also express an abundant S-layer, proteinaceous layers on the surface of the cell that can confer protection from predation, phages and unique surface properties, based on functional annotation and synteny analysis (**Fig. S12**) (*28*). Finally, all Molassodon strains have lost a large island (VPA0713-31; **Fig. 1B**), which encodes, among other things, a putative pilus. Therefore, the differentiation of the

Molassodon ecospecies has proceeded to include many more functions (**Fig. 4H**), some of which also affect motility (**Fig. S7A**), either directly or indirectly.

Discussion

Microbiology has developed powerful technologies to study bacteria in simplified environments, such as agar plates, chemostats and gnotobiotic mice, but assessing the real-world relevance of laboratory phenotypes remains a challenge. Here we show that bacterial ecospecies can be used to elucidate complex bacterial adaptations in the laboratory despite not knowing the natural microhabitats in which those adaptations are used. We have hypothesized that differentiation of ecospecies loci has occurred via adaptation to a new ecological niche (*1*, *8*, *26*). This hypothesis implies that there should be an observable lab phenotype that involves the differentiated genes working together. We have identified and refined a laboratory phenotype for the Molassodon ecospecies, validating the ecological linkage of these genes and allowing us to make a more specific ecological hypothesis, namely that the genes encode an emergent hunt, kill and devour strategy that the bacteria use in viscous liquids.

We first identified a motility phenotype for Molassodon, using transposon mutagenesis to show that a majority of Molassodon genes are involved in enhanced swimming in viscous liquids. Two sets of Molassodon differentiated genes were not necessary for enhanced swimming but were coinduced, namely the Type 6 secretion system and the TonB nutrient uptake system. Based on previous findings in other bacteria linking T6SS effectors and TonB-mediated micronutrient uptake (*29*–*31*), we hypothesized that they might cooperate in killing bacterial prey and devouring them, and were able to generate corresponding phenotypes in our laboratory system. We have not linked these three aspects in a single experiment, and it might be difficult to do so without better mimicking natural hunting grounds but our inference is already supported by combined evidence from coadaptation and coregulation. Understanding TonB coadaptation is likely to be still more challenging, because the two versions we detect are partially shared between ecospecies, and would be aided by determining the specific substrate that each version transports. Currently around a third of ecospecies genes do not have a phenotype, but identifying conditions where they do may help us to further refine our ecological hypothesis and its modelling in the laboratory.

There is a substantial body of work on Type 6-mediated bacterial antagonism and its ecology (*32*), but until recently, it has been hard to confirm that killing was an effective foraging strategy (*5*). We here provide additional experimental and coadaptation-based evidence for this linkage, while additionally highlighting the connection to motility. There has been no direct evidence for active flagella-driven hunting with Type 6 secretion systems (*5*), although mechanisms to actively trap T6SS prey with "grappling hooks" have been reported (*33*). Our experimental and coadaptation results provide evidence for flagella-driven predation, whose effectiveness could be enhanced in a viscous environment by limited nutrient diffusion (*34*). We have also discovered an intermediate genotype which also links nutrition and motility, highlighting their intimate link in this species.

Our results also highlight the key role of introgression of several different functions from closely related species in Molassodon evolution, including flagellin glycosylation and an S-layer. For example, we identified two independent imports of neighboring *lafA, maf2 and glfM* genes, which we functionally linked to further imported genes required to generate the legionaminic acid sugar substrate for glycosylation. As legionaminic acid can also be used to modify LPS independent of *maf* (*22*), this provides a feasible pathway for stepwise assembly of part of the complex phenotype. The distinct (*27*) capsule locus of Molassodon is a reservoir for many other introgressed genes that affect swimming directly or indirectly. Molassodon is the first *Vibrio* species identified with a S-layer, which was acquired late in its evolution and speculatively might provide protection during predation. Our study of Molassodon thus provides evidence for a pan-*Vibrio* gene pool of functionally compatible motility-related genes, in addition to metabolic genes as previously reported (*35*). We propose that this toolbox facilitates continuous evolutionary innovation in dispersal and hunting strategies in the genus, with the resulting diversity of phenotypes explaining the popularity of *Vibrio* amongst researchers studying these functions (*3–5*, *36*, *37*)

We have previously identified ecospecies in three other bacterial species, albeit without functional validation. The phenotypes appear to be completely unrelated but involve facets that are central to the biology and the niche of the species. Putative ecospecies in *Streptococcus mitis* and *S. oralis*, which have convergent evolution in cell-shape related genes (*38*), while *H. pylori* ecospecies have differentiated in genes related to interaction with the immune system and nutrient uptake. Amongst eukaryotes, cichlids are known to speciate rapidly thanks to flexible jaw morphology (*39*) and it could be that speciose bacterial genera also have characteristic axes of innovation underlying their success. Few researchers have searched for ecospecies, but we have recently proposed that such selection-driven structures are highly prevalent amongst bacterial species with high genetic diversity (*38*) and may precipitate speciation (*8*).

Here, we have focused on the genotype-to-phenotype map and evolution of Molassodon, leaving field investigation of the ecology of the ecospecies to future studies. Our approach is thus an example of "reverse ecology", which first uses patterns of gene flow to identify ecologically-relevant groups of strains, before attempting to infer biological and niche differences based on metadata, experiments, and/or functional annotation (*40*, *41*). We have here used ecospecies to extend the approach from recent to complex adaptation without sacrificing sensitivity or specificity in identifying adaptive loci.

Our knowledge of how the phenotypes we have investigated are used in the natural environment is currently extremely limited. We do not know where *V. parahaemolyticus* swarm or use their lateral flagella for swimming. Because Molassodon and Typical strains and intermediate *maf*+ genotypes coexist throughout the geographical range of the species, variation in strain abundances and lateral flagella expression in different microenvironments provides a gateway for investigating the ecological niches associated with swarming and swimming behaviors, as well as their genetic regulation in the natural environment. In the process, we can hope to understand the full ecological context of these fascinating complex adaptations.

Materials and methods

Genomes and sequencing data. Accessions and information for all genomes used for bioinformatics analysis are listed in **Table S1.** Complete genome sequences for some strains were generated using PacBio and Illumina sequencing at Novogene (Beijing, China) and have been deposited at NCBI under Bioproject PRJNA1235929. All other genomes are public and were recovered from NCBI. RNA-seq and Tn-seq data have also been deposited at SRA under Bioproject PRJNA1235929.

To assemble nine new genomes used for phylogenetic analysis (**Table S1**), raw reads were first quality-filtered using Filtlong (v0.2.1) (*42*), followed by genome assembly with Flye (v2.9-b1768) (*43*). Plasmid sequences were removed. A highly curated pangenome was generated for comparison of homologues by RNA-seq and Tn-seq for three Typical and three Molassodon strains: RIMD 2210633 (IPS strain, DFS-0005), SPVP110059 (DFS-0022), VP11007 (DFS-0060), VP09156 (DFS-0007), VP09285 (DFS-0008), and VP10219 (DFS-0010). A high-quality sequence for the six strains was generated using second-generation (Illumina short read) and third-generation (PacBio) sequencing at Novogene (Beijing). High quality genome assemblies were generated using Tricycler version 0.5.3 (*44*). Genomes were annotated using Prokka version 1.14.6 (*45*) and a pangenome was built using Panaroo version 1.3.3 (*46*).

Genome-wide association studies. We conducted GWAS to identify core and accessory genes associated with Molassodon or *maf* genotypes as follows. First, we assembled a set of 1550 non-redundant *V. parahaemolyticus* public genomes (**Table S1**) as follows. We used snp-sites (*47*) to identify SNPs from whole-genome alignments. Next, we calculated the SNP distance matrix and performed hierarchical clustering using the complete linkage method. A threshold of 10,000 SNPs was applied to define clusters, from which we selected the strain with the highest non-redundancy (NR) value within each cluster. To further refine the dataset, we assessed the pairwise SNP distances among the selected strains and removed those with distances below 10,000 SNPs. We then repeated this procedure using the single linkage model to extract an additional set of non-redundant strains. The final dataset comprised the union of non-redundant strains identified through both clustering approaches, ensuring a representative and diverse strain selection. We then used GWES (*8*) to define Molassodon strains (**Fig. S1**). For GWAS based on *maf*, genotype was based on BLAST with an Molassodon *maf2* homolog.

For the SNP-based GWAS, we employed snp-sites (*47*) to extract single nucleotide polymorphisms (SNPs) from whole-genome alignments, generating a variant call format (VCF) file. Association testing was subsequently performed using pyseer (*48*), which evaluates the correlation between SNP variants and the Molassodon phenotype. To further assess genetic differentiation between the groups, we calculated fixation index (FST) values using vcftools (*49*). The predefined groups were used to stratify SNPs for differentiation analysis.

For the gene-based GWAS (presence/absence), we first used panaroo (*46*) to construct a pangenome and extract gene presence-absence variation from whole-genome alignments. To integrate this variation into the GWAS framework, we encoded gene presence-absence patterns as pseudo-SNP loci and subsequently employed snp-sites to generate a VCF file. The resulting VCF file was analyzed using pyseer, employing the same statistical framework as in the SNP-based GWAS. Additionally, gene-level FST values were computed to evaluate differentiation between the Molassodon and Typical (or *maf*+/*maf*- Typical) groups.

For both SNP- and gene-based GWAS, we applied appropriate quality control measures, including a minor allele frequency threshold (> 0.02), to filter out rare variants. The statistical significance of associations was evaluated using the thresholds FST ≥0.05 and -log10(p) ≥10. Differentiated core genes had at least 5 SNPs with FST ≥0.05 and -log10(p) ≥10, while undifferentiated genes were

defined as those with zero SNPs with FST ≥0.05 and -log10(p) ≥10. For the pan-GWAS, the frequency in Molassodon strains was also required to be at least 90%.

Identification of orthologs in *V. parahaemolyticus* and other species in KEGG. The sequences from Molassodon strain UCM-V493 (vph) or Typical *maf*+ strain VP08404 were used for blastn at KEGG and the workflow mafft_default-none-none-fasttree_default as described. Trees were visualized with iTOL (*50*). KEGG genomes are listed in **Table S1**.

Phylogenetic analysis. To generate phylogenetic trees, a set of 130 publicly available genomes plus nine newly-sequenced genomes (see above) were used to build phylogenetic trees, selected to include strains available in our lab for experiments as well as representatives from all groups of interest (*maf* genotype, ecospecies, population) (**Table S1**). To generate the *lafA* (VPA1548 in RIMD2210633) tree, the nucleotide sequence from UCM-V493 (VPUCM_21376) was aligned against the 137 strains using BLASTN. The entire dataset revealed over 82% coverage of the VPA1548 region in the RIMD 2210633 reference genome, which was subsequently used for phylogenetic reconstruction. In total, 124 SNPs were identified in this region and were used to construct a Neighbor-Joining (NJ) tree. Neighbour-joining trees were built using TreeBest and visualized using iTOL (*50*). To generate trees of differentiated/undifferentiated genes, we extracted all SNPs within these regions and constructed a phylogenetic tree using FastTree (*51*).

Additional bioinformatic analysis. *In silico* serotyping was performed using the *V. parahaemolyticus* Kaptive database (*52*, *53*). Synteny analysis was performed with webflags (www.webflags.se) (*54*). Structures of Slr4 homologues and protein complexes (Maf2 + GlfM) were predicted with AlphaFold2 or Aphafold multimer at Colabfold (*23*, *24*). Structures were visualized with Pymol version 3.0.2. RNA-seq coverage was displayed with Integrated Genome Browser (*55*).

Bacterial lab strains and culture conditions. All *V. parahaemolyticus* lab strains were obtained from the Shenzhen Center of Disease Control and Prevention, except for RIMD 2210633 reference isolates, which were kind gifts from Yiquan Zhang (Affiliated Hospital 3 of Nantong University) and Qiyao Wang, East China University of Science and Technology (**Table S6**). A set of sympatric natural strains was isolated from seawater from the coast of Fujian Province, China (平潭, Pingtan and 泉州, Quanzhou) using CHROMagar™Vibrio plates according to the manufacturer's recommendations. *V. parahaemolyticus* strains were routinely grown either on Lysogeny broth (LB) plates or MLB (LB + 3% NaCl) plates at 37°C/30°C, respectively, with 1.5% agar or in broth culture in LB/MLB with shaking at 200 rpm at 37°C/30°C in an aerobic atmosphere. Where appropriate, media was supplemented with marker-selective antibiotics (5/25 μg/ml chloramphenicol (Cm) for *V. parahaemolyticus*/*E. coli*, 50 μg/ml kanamycin (Kan)). *E. coli* strains (**Table S6**) were grown aerobically at 37°C in LB broth or on LB agar supplemented with the appropriate antibiotics for marker selection as listed above.

Motility assays. Strains were grown in MLB until log phase, and then 0.05 OD_{600} was inoculated into swim or swarm plates. Swim plates were made from 0.25-0.325% agar as described previously (*56*). For swarming, we used either previously described BHI agar (1.5%) (*56*), or for Tn-seq, LB agar with 2% NaCl. Plates were incubated at 30°C for approximately 16 hours until halos were observed. For diameter measurements, if the colony was not round, which means the longest diameter is more than 1 mm longer than the shortest diameter, we took the average between the longest and the shortest diameters.

General recombinant DNA techniques. All plasmids generated and/or used in this study are listed in **Table S13**. Oligonucleotide primers (Sangon) are listed in **Table S14**. DNA constructs and mutations were confirmed by Sanger sequencing (Sangon). Restriction enzymes, *Taq* polymerase for validation PCR, and T4 DNA ligase were purchased from Tiangen, ThermoFisher Scientific, or New

England Biologs. For cloning purposes, Phanta high-fidelity DNA polymerase was used (Vazyme). Some plasmids were generated with the Sangon Seamless cloning kit.

***V. parahaemolyticus* deletion mutant construction.** *V. parahaemolyticus* deletion mutant strains were generated via recombination with suicide plasmids generated based on pDM4 (*57*), as described previously (*56*, *58*). As an example, the deletion of *lafK* in RIMD 2210633 is described. Oligonucleotides and plasmids used for other mutants are listed in **Tables S13** & **S14**. Approximately 1kb upstream/downstream of the region to be deleted was amplified with primers DFO-0020/0021 and DFO-0022/0023 from WT genomic DNA (DFS-0191), respectively. This introduced a short (approximately 25 bp) region of homology between the two 1kb fragments for overlap PCR, as well as *Xba*I and *Xho*I restriction sites for cloning into pDM4. The upstream and downstream fragments were mixed in an equimolar ratio and subjected to overlap PCR using DFO-0020/0023, and pDM4 was amplified with DFO-0018/0019 and digested with *Dpn*I to remove template DNA. The overlap PCR product and pDM4 were digested with *Xba*I/*Xho*I and then ligated overnight with T4 DNA ligase.

In some cases, the Sangon seamless cloning kit was used. Here, PCR primers for upstream and downstream homology regions also introduced regions overlapping pDM4 (amplified with DFO-0412/0413). For example, to generate a *lafK* deletion construct for Molassodon strains, fragments were amplified with DFO-0410/0281 and DFO-0282/0411. Fragments were mixed in approximately equimolar ratios and fused after addition of seamless cloning mix for 30 min at 50°C.

Ligations/seamless cloning reactions were transformed into DH5α λ*pir* and Cm^R colonies were verified by colony PCR with DFO-0044/0045. Verified plasmids (*e.g.*, pSSvSH3 for *lafK*) were transformed into *E. coli* SM10 λ*pir* or *E. coli* MFD λ*pir,* followed by conjugation into the *appropriate V. parahaemolyticus* recipient strain. Transconjugants with single crossovers were first selected on MMM (minimal marine medium) plates with 10 μg/ml Cm. Single colonies were restreaked to purity and then transferred to MMM plates containing 15% sucrose to counter-select for double-crossover strains. Deletion clones were verified by colony PCR with DFO-0024/0023 (for Δ*lafK*) and checked for Cm sensitivity to ensure the second recombination had occurred.

Expression of *maf2-glfM* in Typical *V. parahaemolyticus.* To overexpress Molassodon *maf2-glfM* in Typical strain, the two genes with their native RBS were amplified and short, overlapping sequences were added by PCR using DFO-0942/0943 from DFS-0010 gDNA, the pMMB207 plasmid skeleton carrying the Ptac promoter was linearized by PCR with DFO-0437/0438. The two linear DNA fragments were then assembled into plasmids by ligation-independent cloning (CPEC, circular polymerase extension cloning) (*59*) and transformed into DH5α λ*pir*. The resulting plasmid, pWHQ222, was verified by colony PCR with DFO-0306/0187 and sequencing with DFO-0168, and then conjugated from *E. coli* MFD λpir into the appropriate recipient *V. parahaemolyticus* strain.

sfGFP reporter construction. To generate a transcriptional reporter for the Molassodon *lafA* (DFS0010_4645) promoter, approximately 120 bp upstream of the *lafA* start codon was amplified with DFO-0300/0302 from DFS-0007 gDNA to add an *Xba*I site at the 5' end and a region overlapping the strong non-native ribosome binding site (RBS) and GFP from pXG10-SF (*60*) at the 3' end. The superfolder GFP (sfGFP) gene was amplified with the RBS from pXG10-SF with DFO-0180/0139, which introduced a *Pst*I site at the 3' end of sfGFP. The two fragments were fused by overlap-extension PCR with DFO-0300/0139. The conjugatable replicating plasmid pMMB207 (*61*) was amplified by inverse PCR with DFO-0279/0185 to add *Xba*I and *Pst*I sites and digested with *Dpn*I to remove template DNA. The vector and overlap PCR fragments were then digested with *Xba*I and *Pst*I, ligated, and transformed into DH5α λ*pir*. The resulting plasmid, pSSvSH47, was verified by colony PCR with DFO-0306/0187 and sequencing with DFO-0168, and then conjugated from SM10 λ*pir* into the appropriate recipient *V. parahaemolyticus* strain.

To generate the corresponding reporter for RIMD2210633 P*lafA*, CPEC was employed to replace the Molassodon *lafA* promoter in pSSvSH47. Approximately 170 bp upstream of the *lafA* start codon with short, overlapping sequences was amplified with DFO-0384/0443 from RIMD2210633 gDNA and a GFP reporter skeleton was amplified with DFO-0442/0385 from pSSvSH47. The two linear DNA fragments were then assembled into plasmids using a Ready-to-Use Seamless Cloning Kit (Sangon) and transformed into DH5α λ*pir*. The resulting plasmid, pWHQ201, was verified by colony PCR with DFO-0306/0187 and sequencing with DFO-0168, and then conjugated from SM10 λ*pir* into the appropriate recipient *V. parahaemolyticus* strain.

P*lafA* reporter assays. To measure the P*lafA* reporters, strains were grown in MLB with or without 3% (w/v) PVP (polyvinylpyrrolidone K90, Xushuo Biotechnology Co., Ltd., Shanghai, China) at 30°C, and then optical density (OD_{600}) and GFP (excitation 485 nm/emission 528 nm), were measured every 15 minutes.

RNA isolation and cDNA library preparation for RNA-seq. Bacterial cells from cultures or motility plates were mixed with 0.2 volumes stop mix (5% phenol, 95% ethanol) and immediately frozen at -80°C. RNA was extracted from thawed cell pellets for total RNA using hot phenol, respectively, as described previously (*62*). Total RNA was digested with DNase I, depleted of rRNA, and converted to cDNA libraries at Novogene, Inc (Beijing, China).

Processing of RNA-seq data and RNA-seq expression analysis. Fastp version 0.23.4 was used for quality control of raw reads with default parameters (*63*). Reads were then mapped to the corresponding reference genome using Segemehl version 0.3.4 with default parameters (*64*). FeatureCounts version 2.0.3 was used for gene expression quantification with parameters "-B -C -f -p -s 1 -O --fraction --countReadPairs --ignoreDup –primary" (*65*). Differential expression was performed using R package DESeq2 version 1.42.1 using the output from featureCounts without normalization (*66*). Gene set enrichment analysis was performed using R package clusterProfiler version 4.10.1 on differentially expressed genes passing a filter of adjusted P-value < 0.05 (*67*). Gene set classifications include GO terms (*68*), KEGG pathways, KEGG KO (*69*), COG categories (*70*), and Molassodon genes based on GWAS. Other general sequencing file and alignment file manipulations (*e.g.*, read reverse complementation, alignment statistics, genome coverage file generations, etc.) and calculations (e.g. TPM normalization) were performed using Samtools version 1.19 (*71*) and Julia. Plots were generated using Julia package Makie.jl (*72*).

Tn-seq data generation. To generate a high-density Tn pool, we used a Mariner transposon delivered by conjugation on a suicide (R6K origin) plasmid. We modified the plasmid pCAT-Mariner(*73*) to be compatible with current Illumina platforms as follows. Specifically, Illumina sequencing components were added to both sides of the Mariner transposon repeat region through PCR using primers DFO-690/672, DFO-690/0673, DFO-0689/0684, and DFO-689/683 from pCAT-Mariner. The four resulting linear DNA fragments were then assembled into plasmids CPEC and transformed into DH5α λpir. The resulting plasmid, pWHQ011, was verified by sequencing with primers DFO-672 and DFO-673 for each side of the repeat region.

Following conjugation with *E. coli* MFD λ*pir*, putative Tn mutants were selected on LB with Cm at 10 μg/ml. Approximately ~500,000/~800,000 colonies (Typical/RIMD 2210633/DFS-0191 and Molassodon/DFS-0010) were recovered and immediately frozen at -80°C with 25% glycerol. After sequencing (see next section) we detected 168,777 for the RIMD 2210633 pool and 331,277 unique insertions for the DFS-0010 pool.

Motility selection was applied as follows. For swarming (RIMD 2210633 only) cells were recovered from the -80°C freezer by shaking in MLB Broth with Cm at 10 μg/ml for 3 hours at 30°C. Cells were then resuspended in MLB and approximately 10^7 CFU in 10 μl was spotted on 2% MLB plates, which we

found generated robust swarming, 10 times. After approximately 16 hours of incubation at 30°C, we harvested the outermost edge of swarm halos (Swarm-out), as well as the inside (Swarm-in) as a reference. For swimming (both Typical and Molassodon), cells were recovered from the -80°C freezer by shaking in MLB Broth with Cm at 10 μg/ml for 3 hours at 30°C. Next, approximately 10^8 CFU in 100 μl was inoculated into swim agar (*56*) in a large beaker to ensure enough material could be recovered for library preparation. Assays were incubated at room temperature for 16 hours. As a reference, we inoculated the same swim agar with the mutant pool and shook the tube at room temperature during the same time period to remove the selection for motility, maintain the same nutrient conditions, and obtain enough material for sequencing. To isolate mutants from the swimming agar, the swimming agar containing mutants was collected and centrifuged at 3500 rpm for 10 minutes to precipitate the bacterial cells. Subsequently, the bacteria were repeatedly washed with an equal volume of PBS until the agar was completely removed from the precipitate.

Tn-seq libraries were generated from input (original mutant pools, no selection), reference, and motility libraries, essentially as described previously (*74*), as follows. Genomic DNA was extracted using the FastPure Blood/Cell/Tissue/Bacteria DNA Isolation Mini Kit (Vazyme). Subsequently, 5 μg of DNA was digested with *Mme*I (New England Biolabs), and transposon-sized fragments of approximately 1500 bp were extracted via gel extraction (Gel DNA Recovery Kit, Zymo) after separation on a 2% agarose gel. Annealed sequencing adapters were ligated overnight at room temperature. The ligated samples were then used as templates for final library PCR enrichment, which was performed for 22 cycles using an adapter-specific primer (DFO-0762) and a transposon-specific primer containing different Illumina indices (DFO-0936 - DFO-0941). The resulting 171 bp products were purified by gel extraction and sequenced on the Illumina NovaSeq 6000 system.

Tn-seq data analysis. Raw data were first subjected to quality control using fastP (v0.23.4) and processed with cutadapt (v4.8) to remove sequencing adaptors and transposon sequences. The remaining 14 - 16 bp transposon-junction sequences were mapped to RIMD 2210633 or DFS-0010 chromosomes using Bowtie (v1.3.1) (*75*). Subsequently, the read counts of each insertion site were tallied using the TraDIS toolkit (*76*). In order to compare the changes in the relative abundance of each mutant gene across different samples, the total read counts of each sample were normalized to the lowest count value in each group. Following normalization, the log - ratio of output read counts to reference read counts was calculated for each gene, serving as a crucial metric to assess the motility contribution of that particular gene.

Transmission electron microscopy. For visualization of lateral flagella, the appropriate strains were grown in MLB liquid culture or on MLB plates. For liquid cultures, 100 μl of overnight culture was harvested, gently washed twice with 1x PBS, and resuspended in 1x PBS to a final OD600 of approximately 0.1. For samples from agar plates, we resuspended part of the colony in 1 ml of 1x PBS. A 300 mesh copper grid (EMCN) was prepared by glow discharge (PELCO easiGlow 91000), and 10 μl of washed cells was added. After 1 min, liquid was removed gently with a filter paper and the grid was washed with 10 μl 1xl PBS. The grid was then negative-stained with uranyl acetate for 20 seconds. The uranyl acetate was removed with a filter paper. The grid was quickly washed once with sdH_2O and dried under a heat lamp.

Type VI killing assay. T6SS2 killing activity was measured against *E. coli* BW25113 (DFS-2143) carrying a Kan^R plasmid (pWHQRed2) to allow plating for colony-forming units (CFUs). To generate the plasmid, approximately 450 bp upstream of the *gyrB* start codon with short, overlapping sequences was amplified with DFO-1078/1072 from RIMD 2210633 gDNA, the RFP gene was amplified with its RBS from pRP0122-Pbe-AmCyan_Bc3-5_turboRFP with DFO-1073/1087 and a plasmid backbone was amplified with DFO-1088/1081 from pET28a. The three linear DNA fragments were then assembled into plasmids using a Ready-to-Use Seamless Cloning Kit (Sangon) and

transformed into DH5α λ*pir*. The resulting plasmid, pWHQRed2, was verified by colony PCR with DFO-1078/0187 and sequencing with DFO-0187, and then electroporated into the *E. coli* BW25113 strain. Predator strains included Typical and Molassodon wild-type, Δ*hcp2*, and Δ*lafA* strains (**Table S6**). To initiate the killing assay, overnight cultures of prey and predator strains were washed with PBS one time and resuspended by pipetting. Predator and prey suspensions were diluted to a final OD_{600} of 0.5 with M9 broth with 2% NaCl and 5% PVP. Predator and prey strains were mixed at a ratio of 10:1 and incubated at 30°C. A control with *E. coli* only was included. After 3.5 hours, serial dilutions were spotted onto antibiotic-containing LB agar plates to enumerate prey colony forming units (CFU/ml). Data were normalized using the CFU of the corresponding single *E. coli* strain group.

Growth assays. First, a whole cell lysate of prey *E. coli* BW25113 was generated as follows. Approximately 60 OD_{600} of cells were harvested from LB broth cultures grown at 3°C until log phase and washed three times with PBS and lysed by sonication. The sonicated lysate was filtered through a 0.22 μm filter (Sangon). *V. parahaemolyticus* strains were grown in pre-culture in 2 ml MLB at 30°C until approximately log phase, washed three times with 1× PBS, and inoculated into test media at a final OD600 of 0.5. Each strain was grown in triplicate in 200 μl of MLB or M9 (salts only, no carbon source) with ⅔ volume *E. coli* lysate with shaking at 30°C. Results shown are representative of two independent experiments.

Statistical analysis. Statistical tests were performed using Graphpad prism.

Acknowledgements

We thank Qiyao Wang, Yiquan Zhang, Kim Orth, and Melanie Blokesch for providing strains and advice. **Funding:** SLS is supported by an RFIS-II award from the National Natural Science Foundation of China (NSFC) (#32250610209) and DFG (German Research Foundation) Start-up funding within the framework of Priority Program SPP2002. Work in the DF laboratory is supported by an NSFC RFIS-III award (#32350710791). CY is supported by the National Natural Science Foundation of China (No. 32270003), and YC is supported by the Natural Science Foundation of China (92478118, 32270064) and Shanghai Municipal Science and Technology Commission (24ZR1493200). **Author contributions:** Performed experiments: SLS, ZJ, HW, ZX. Bioinformatic analysis: SLS, ZJ, HW, HF, WX, CY. Designed research: SLS, DF. Materials and strains: YC. Analysis and interpretation: SLS, DF. Manuscript writing: SLS, DF. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Newly-sequenced genomes, RNA-seq and Tn-seq data have been deposited at SRA under Bioproject PRJNA1235929.

References

1. E. Tourrette, R. C. Torres, S. L. Svensson, T. Matsumoto, M. Miftahussurur, K. A. Fauzia, R. I. Alfaray, R.-K. Vilaichone, V. P. Tuan, *Helicobacter* Genomics Consortium, D. Wang, A. Yadegar, L. M. Olsson, Z. Zhou, Y. Yamaoka, K. Thorell, D. Falush, An ancient ecospecies of *Helicobacter pylori*. *Nature* (2024), doi:10.1038/s41586-024-07991-z.
2. C. A. Broberg, T. J. Calder, K. Orth, *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes Infect.* **13**, 992–1001 (2011).
3. M. Grognot, J. W. Nam, L. E. Elson, K. M. Taute, Physiological adaptation in flagellar architecture improves *Vibrio alginolyticus* chemotaxis in complex environments. *Proc Natl Acad Sci USA*. **120**, e2301873120 (2023).

4. L. Speare, A. G. Cecere, K. R. Guckes, S. Smith, M. S. Wollenberg, M. J. Mandel, T. Miyashiro, A. N. Septer, Bacterial symbionts use a Type VI secretion system to eliminate competitors in their natural host. *Proc Natl Acad Sci USA*. **115**, E8528–E8537 (2018).
5. A. K. M. Stubbusch, F. J. Peaudecerf, K. S. Lee, L. Paoli, J. Schwartzman, R. Stocker, M. Basler, O. T. Schubert, M. Ackermann, C. Magnabosco, G. G. D'Souza, Antagonism as a foraging strategy in microbial communities. *Science*. **388**, 1214–1217 (2025).
6. C. Freitas, T. Glatter, S. Ringgaard, The release of a distinct cell type from swarm colonies facilitates dissemination of *Vibrio parahaemolyticus* in the environment. *ISME J.* **14**, 230–244 (2020).
7. M. Homma, T. Nishikino, S. Kojima, Achievements in bacterial flagellar research with focus on *Vibrio* species. *Microbiol. Immunol.* **66**, 75–95 (2022).
8. Y. Cui, C. Yang, H. Qiu, H. Wang, R. Yang, D. Falush, The landscape of coadaptation in *Vibrio parahaemolyticus. eLife*. **9** (2020), doi:10.7554/eLife.54136.
9. C. Yang, X. Pei, Y. Wu, L. Yan, Y. Yan, Y. Song, N. M. Coyle, J. Martinez-Urtaza, C. Quince, Q. Hu, M. Jiang, E. Feil, D. Yang, Y. Song, D. Zhou, R. Yang, D. Falush, Y. Cui, Recent mixing of *Vibrio parahaemolyticus* populations. *ISME J.* **13**, 2578–2588 (2019).
10. Y. Chen, J. Dai, J. G. Morris, J. A. Johnson, Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol.* **10**, 274 (2010).
11. L. McCarter, M. Hilmen, M. Silverman, Flagellar dynamometer controls swarmer cell differentiation of *V. parahaemolyticus. Cell*. **54**, 345–351 (1988).
12. L. McCarter, M. Silverman, Surface-induced swarmer cell differentiation of *Vibrio parahaemolyticus*. *Mol. Microbiol.* **4**, 1057–1062 (1990).
13. C. J. Gode-Potratz, R. J. Kustusch, P. J. Breheny, D. S. Weiss, L. L. McCarter, Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. *Mol. Microbiol.* **79**, 240–263 (2011).
14. I. Kawagishi, M. Imagawa, Y. Imae, L. McCarter, M. Homma, The sodium-driven polar flagellar motor of marine *Vibrio* as the mechanosensor that regulates lateral flagellar expression. *Mol. Microbiol.* **20**, 693–699 (1996).
15. A. N. Septer, G. Sharpe, E. A. Shook, The *Vibrio fischeri* type VI secretion system incurs a fitness cost under host-like conditions. *BioRxiv* (2023), doi:10.1101/2023.03.07.529561.
16. L. Tang, S. Dong, N. Rasheed, H. W. Wu, N. Zhou, H. Li, M. Wang, J. Zheng, J. He, W. C. H. Chao, *Vibrio parahaemolyticus* prey targeting requires autoproteolysis-triggered dimerization of the type VI secretion system effector RhsP. *Cell Rep.* **41**, 111732 (2022).
17. S. C. Booth, O. J. Meacock, K. R. Foster, Cell motility empowers bacterial contact weapons. *ISME J.* **18** (2024), doi:10.1093/ismejo/wrae141.
18. R. Gallegos-Monterrosa, S. J. Coulthurst, The ecological impact of a bacterial weapon: microbial interactions and the Type VI secretion system. *FEMS Microbiol. Rev.* **45** (2021), doi:10.1093/femsre/fuab033.
19. B. T. Ho, T. G. Dong, J. J. Mekalanos, A view to a kill: the bacterial type VI secretion system. *Cell Host Microbe*. **15**, 9–21 (2014).
20. R. J. Kustusch, C. J. Kuehl, J. H. Crosa, The *ttpC* gene is contained in two of three TonB systems in the human pathogen *Vibrio vulnificus*, but only one is active in iron transport and virulence. *J. Bacteriol.* **194**, 3250–3259 (2012).
21. J. Unay, N. Kint, P. H. Viollier, Evolution of paralogous multicomponent systems for site-specific O-sialylation of flagellin in Gram-negative and Gram-positive bacteria. *Curr. Biol.* **34**, 2932-2947.e7 (2024).
22. N. Hashii, Y. Isshiki, T. Iguchi, S. Kondo, Structural analysis of the carbohydrate backbone of *Vibrio parahaemolyticus* O2 lipopolysaccharides. *Carbohydr. Res.* **338**, 1063–1071 (2003).
23. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat. Methods*. **19**, 679–682 (2022).
24. R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. W. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K.

Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, D. Hassabis, Protein complex prediction with AlphaFold-Multimer. *BioRxiv* (2021), doi:10.1101/2021.10.04.463034.

25. M. J. Morrison, B. Imperiali, The renaissance of bacillosamine and its derivatives: pathway characterization and implications in pathogenicity. *Biochemistry*. **53**, 624–638 (2014).
26. Y. Cui, X. Yang, X. Didelot, C. Guo, D. Li, Y. Yan, Y. Zhang, Y. Yuan, H. Yang, J. Wang, J. Wang, Y. Song, D. Zhou, D. Falush, R. Yang, Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen *Vibrio parahaemolyticus*. *Mol. Biol. Evol.* **32**, 1396–1410 (2015).
27. S. Bian, W. Zeng, Q. Li, Y. Li, N.-K. Wong, M. Jiang, L. Zuo, Q. Hu, L. Li, Genetic Structure, Function, and Evolution of Capsule Biosynthesis Loci in *Vibrio parahaemolyticus*. *Front. Microbiol.* **11**, 546150 (2020).
28. S. Ali, B. Jenkins, J. Cheng, B. Lobb, X. Wei, S. Egan, T. C. Charles, B. J. McConkey, J. Austin, A. C. Doxey, Slr4, a newly identified S-layer protein from marine Gammaproteobacteria, is a major biofilm matrix component. *Mol. Microbiol.* **114**, 979–990 (2020).
29. M. Si, C. Zhao, B. Burkinshaw, B. Zhang, D. Wei, Y. Wang, T. G. Dong, X. Shen, Manganese scavenging and oxidative stress response mediated by type VI secretion system in *Burkholderia thailandensis*. *Proc Natl Acad Sci USA*. **114**, E2233–E2242 (2017).
30. M. Si, Y. Wang, B. Zhang, C. Zhao, Y. Kang, H. Bai, D. Wei, L. Zhu, L. Zhang, T. G. Dong, X. Shen, The Type VI Secretion System Engages a Redox-Regulated Dual-Functional Heme Transporter for Zinc Acquisition. *Cell Rep.* **20**, 949–959 (2017).
31. T. Wang, M. Si, Y. Song, W. Zhu, F. Gao, Y. Wang, L. Zhang, W. Zhang, G. Wei, Z.-Q. Luo, X. Shen, Type VI secretion system transports Zn^{2+} to combat multiple stresses and host immunity. *PLoS Pathog.* **11**, e1005020 (2015).
32. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, The evolution and ecology of bacterial warfare. *Curr. Biol.* **29**, R521–R537 (2019).
33. Y.-W. Lien, D. Amendola, K. S. Lee, N. Bartlau, J. Xu, G. Furusawa, M. F. Polz, R. Stocker, G. L. Weiss, M. Pilhofer, Mechanism of bacterial predation via ixotrophy. *Science*. **386**, eadp0614 (2024).
34. C. D. Nadell, C. J. Marx, Bacteria poison and eat their neighbors. *Science*. **388**, 1134–1135 (2025).
35. J.-H. Hehemann, P. Arevalo, M. S. Datta, X. Yu, C. H. Corzett, A. Henschel, S. P. Preheim, S. Timberlake, E. J. Alm, M. F. Polz, Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat. Commun.* **7**, 12860 (2016).
36. W. Zhao, F. Caro, W. Robins, J. J. Mekalanos, Antagonism toward the intestinal microbiota and its effect on *Vibrio cholerae* virulence. *Science*. **359**, 210–213 (2018).
37. S. B. Otto, R. Servajean, A. Lemopoulos, A.-F. Bitbol, M. Blokesch, Interactions between pili affect the outcome of bacterial competition driven by the type VI secretion system. *Curr. Biol.* **34**, 2403-2417.e9 (2024).
38. C. Yang, H. Huang, N. Wang, X. Didelot, R. Yang, Y. Cui, D. Falush, Macrogenetic atlas of prokaryotes. *BioRxiv* (2025), doi:10.1101/2025.02.26.640246.
39. N. Peoples, M. D. Burns, M. Mihalitsis, P. C. Wainwright, Evolutionary lability of a key innovation spurs rapid diversification. *Nature* (2025), doi:10.1038/s41586-025-08612-z.
40. P. Arevalo, D. VanInsberghe, J. Elsherbini, J. Gore, M. F. Polz, A reverse ecology approach based on a biological definition of microbial populations. *Cell*. **178**, 820-834.e14 (2019).
41. B. J. Shapiro, M. F. Polz, Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol.* **22**, 235–247 (2014).
42. GitHub - rrwick/Filtlong: quality filtering tool for long reads, (available at https://github.com/rrwick/Filtlong).
43. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
44. R. R. Wick, L. M. Judd, L. T. Cerdeira, J. Hawkey, G. Méric, B. Vezina, K. L. Wyres, K. E. Holt, Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* **22**, 266 (2021).
45. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. **30**, 2068–2069 (2014).

46. G. Tonkin-Hill, N. MacAlasdair, C. Ruis, A. Weimann, G. Horesh, J. A. Lees, R. A. Gladstone, S. Lo, C. Beaudoin, R. A. Floto, S. D. W. Frost, J. Corander, S. D. Bentley, J. Parkhill, Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
47. A. J. Page, B. Taylor, A. J. Delaney, J. Soares, T. Seemann, J. A. Keane, S. R. Harris, SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2**, e000056 (2016).
48. J. A. Lees, M. Galardini, S. D. Bentley, J. N. Weiser, J. Corander, pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics*. **34**, 4310–4312 (2018).
49. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).
50. I. Letunic, P. Bork, Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
51. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
52. R. R. Wick, E. Heinz, K. E. Holt, K. L. Wyres, Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. *J. Clin. Microbiol.* **56** (2018), doi:10.1128/JCM.00197-18.
53. L. van der Graaf-van Bloois, H. Chen, J. A. Wagenaar, A. L. Zomer, Development of Kaptive databases for *Vibrio parahaemolyticus* O- and K-antigen genotyping. *Microb. Genom.* **9** (2023), doi:10.1099/mgen.0.001007.
54. C. K. Saha, R. Sanches Pires, H. Brolin, M. Delannoy, G. C. Atkinson, FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics*. **37**, 1312–1314 (2021).
55. N. H. Freese, D. C. Norris, A. E. Loraine, Integrated genome browser: visual analytics platform for genomics. *Bioinformatics*. **32**, 2089–2095 (2016).
56. S. Chimalapati, A. E. Lafrance, L. Chen, K. Orth, *Vibrio parahaemolyticus*: Basic Techniques for Growth, Genetic Manipulation, and Analysis of Virulence Factors. *Curr. Protoc. Microbiol.* **59**, e131 (2020).
57. X. Mou, E. J. Spinard, M. V. Driscoll, W. Zhao, D. R. Nelson, H-NS is a negative regulator of the two hemolysin/cytotoxin gene clusters in Vibrio anguillarum. *Infect. Immun.* **81**, 3566–3576 (2013).
58. C. Yang, H. Qiu, S. L. Svensson, C. Ni, S. Gao, Z. Jia, H. Wen, L. Xie, W. Xu, Y. Qin, S. Lin, J. Wang, Y. Zhang, Y. Li, M. Jiang, X. Shi, Q. Hu, Z. Zhou, Y. Chao, R. Yang, D. Falush, Pandemic *Vibrio parahaemolyticus* wave succession by gene loss. *BioRxiv* (2024), doi:10.1101/2024.06.03.596356.
59. J. Quan, J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE*. **4**, e6441 (2009).
60. C. P. Corcoran, D. Podkaminski, K. Papenfort, J. H. Urban, J. C. D. Hinton, J. Vogel, Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.* **84**, 428–445 (2012).
61. V. M. Morales, A. Bäckman, M. Bagdasarian, A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene*. **97**, 39–47 (1991).
62. C. M. Sharma, F. Darfeuille, T. H. Plantinga, J. Vogel, A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* **21**, 2804–2817 (2007).
63. S. Chen, Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. **2**, e107 (2023).
64. S. Hoffmann, C. Otto, G. Doose, A. Tanzer, D. Langenberger, S. Christ, M. Kunz, L. M. Holdt, D. Teupser, J. Hackermüller, P. F. Stadler, A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.* **15**, R34 (2014).
65. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

67. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. **2**, 100141 (2021).
68. Gene Ontology Consortium, S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas, K. Van Auken, M. Westerfield, The Gene Ontology knowledgebase in 2023. *Genetics*. **224**, iyad031 (2023).
69. M. Kanehisa, Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
70. M. Y. Galperin, Y. I. Wolf, K. S. Makarova, R. Vera Alvarez, D. Landsman, E. V. Koonin, COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2021).
71. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience*. **10** (2021), doi:10.1093/gigascience/giab008.
72. S. Danisch, J. Krumbiegel, Makie.jl: Flexible high-performance data visualization for Julia. *JOSS*. **6**, 3349 (2021).
73. Z. Wang, S. Zhu, C. Li, L. Lyu, J. Yu, D. Wang, Z. Xu, J. Ni, B. Gao, J. Lu, Y.-F. Yao, Gene essentiality profiling reveals a novel determinant of stresses preventing protein aggregation in *Salmonella*. *Emerg. Microbes Infect.* **11**, 1554–1571 (2022).
74. A. L. Goodman, N. P. McNulty, Y. Zhao, D. Leip, R. D. Mitra, C. A. Lozupone, R. Knight, J. I. Gordon, Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe*. **6**, 279–289 (2009).
75. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
76. L. Barquist, M. Mayho, C. Cummins, A. K. Cain, C. J. Boinett, A. J. Page, G. C. Langridge, M. A. Quail, J. A. Keane, J. Parkhill, The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics*. **32**, 1109–1111 (2016).