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. Coadaptation within the "Molassodon" ecospecies of *Vibrio parahaemolyticus* is concentrated in 48 core genes and over 100 accessory genes on an otherwise undifferentiated genetic background. Using this list of adaptive differences as a foundation for laboratory experiments, we find that the ecological strategy of the ecospecies 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 VI 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 challenge.

Complex adaptations such as the flight of birds, the bacterial flagellum or the feeding strategies of orcas [1] 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. One focus for investigation has been evolution that happens repeatedly, such as marine sticklebacks colonising rivers. Within each river, similar combinations of alleles conferring freshwater-adaptive traits are assembled onto otherwise unrelated genetic backgrounds, making it possible to identify regions of the genome responsible for adaptation [2].

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 [3]. We have proposed that the differences between ecospecies is generated and maintained by natural selection. High rates of homologous recombination mean that the ecospecies genotype is found on diverse genetic backgrounds. Consequently, the analysis tools used for sticklebacks and other eukaryotic ecological genomics model organisms can be applied to identify adaptation, which can generate a list of differentiated genome regions that can be used for further functional genomic investigation [2].

For example, *Helicobacter pylori* colonizes the stomachs of humans and occasionally other mammals. The Hardy ecospecies is differentiated from typical Ubiquitous *H. pylori* at 100 of approximately 1500 genes and a handful of accessory genes [3]. 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. Rampant recombination during mixed infections mean that traces of clonal descent within *H. pylori* genomes disappear within a few hundred years [4], but characteristic differences between Hardy and Ubiquitous strains have been maintained for at least three hundred thousand years and across diverse *H. pylori* ancestries [3], implying stable strong selection to maintain ecospecies differentiation at each site, in several human populations.

Some strains of *Vibrio parahaemolyticus* can also briefly colonize humans, leading to gastroenteritis [5]. 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 VI secretion-mediated bacterial killing [6–9]. We previously identified a highly distinct group of *V. parahaemolyticus*, EG1a, characterized by differentiation at genes including lateral flagella [10], secondary flagella usually used for swarming motility on surfaces by all *V. parahaemolyticus* and a small subset of other *Vibrio* spp. [11]. 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. Additional EG1a 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 rename the ecospecies "Molassodon", evoking hunting in sticky molasses, with strains lacking the genotype designated as Typical. 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

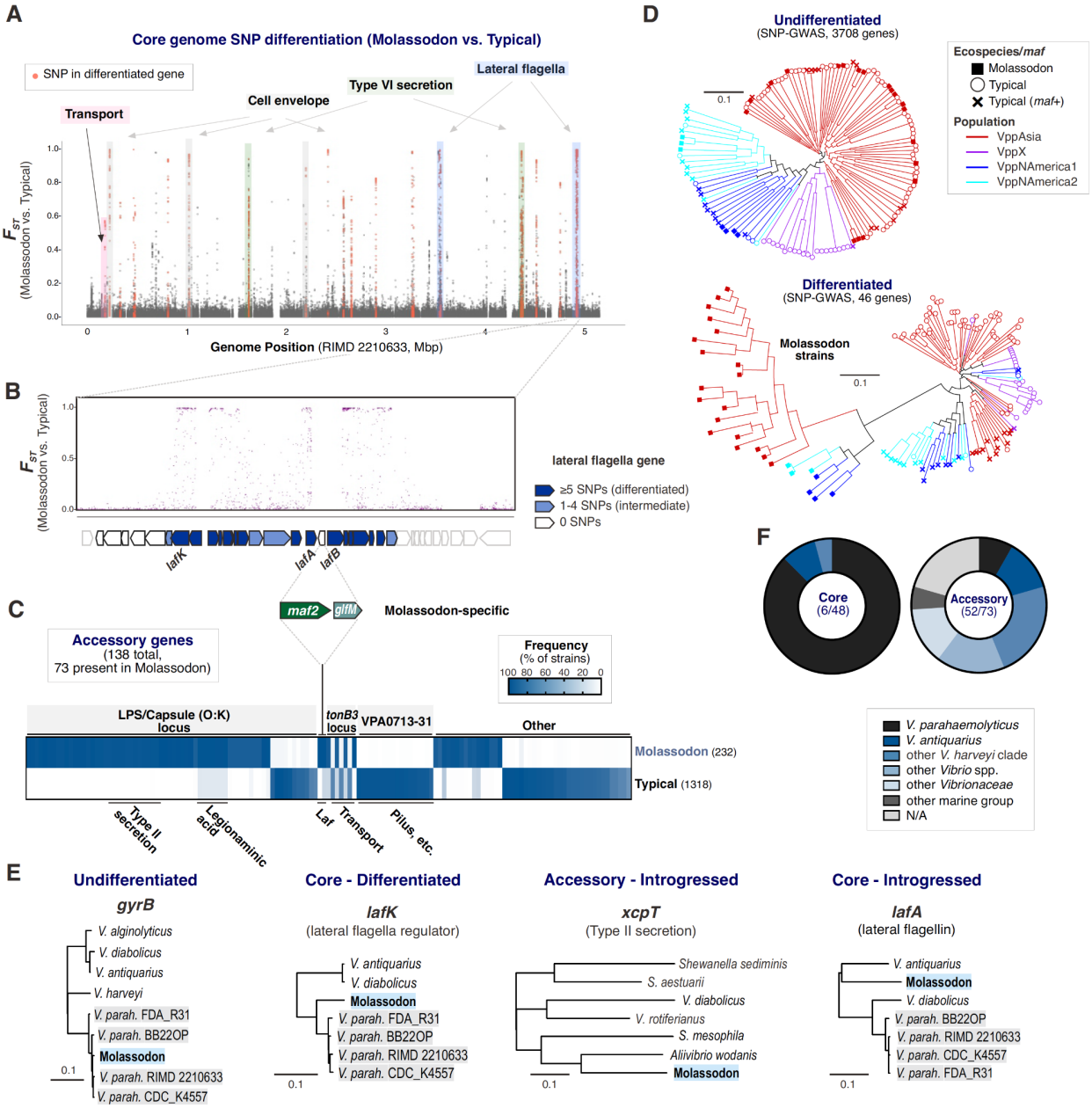

Figure 1. The Molassodon ecospecies genotype. (A) Manhattan plot for core genome SNP (single-nucleotide polymorphism) GWAS of Molassodon (n=223) vs. Typical (n=1318) strains. Significant SNPs: F_{st} (fixation index) ≥0.5 and -log10(p) ≥10. Only those in differentiated genes (5 or more SNPs) are labeled in red. Key functions relevant to this study are highlighted. See also **Tables S2-S4. (B)** Zoom-in on indicated lateral flagella region in F_{sy} plot in panel A**. C)** Frequency of genes identified as significantly differentially distributed between Molassodon and Typical strains via pangenome GWAS. See also **Table S5**. Significant genes: F_{st} ≥ 0.5 and -$\log_{10}$(p) ≥ 10, also requiring >90% or <10% frequency in Molassodon. Genes are ordered by putative function. **(D)** Approximate Maximum Likelihood g 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 F_{st} ≥0.05. Note that two highly differentiated core genes identified in the pan-GWAS are not included. Branches: geographic population. Tips: ecospecies

(Typical/Molassodon) and *maf* (*maf2-glfM* accessory genes) genotype. See also **Table S3** for Differentiated and Undifferentiated genes. **(E)** Representative trees for KEGG matches for undifferentiated (*gyrB*) and representative example genes for the three types of Molassodon-differentiated genes: 1) Core genes with closest match in other *V. parahaemolyticus* strains, but with an extended branch length; 2) Accessory genes with signs of introgression (*i.e.*, closest match outside of *V. parahaemolyticus*); 3) Core genes with signs of introgression. Grey: Typical *V. parahaemolyticus*. Blue: Molassodon. Pink: Best species match in the database. For additional examples, see **Fig. S5**. **(F)** Taxon of best KEGG ortholog (by % amino acid identity) for Molassodon-present accessory and core genes. N/A: no match above 40%.

clustering of a large group of interacting genes [10]. 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 [12] to identify SNPs (single-nucleotide polymorphisms) differentiating Molassodon strains from Typical strains (**Fig. 1A; Tables S1-S4; Figs. S1A-S1C**). Differentiation adjacent to highly differentiated genes decays to the low levels found in most of the genome within a couple of hundred bases (*e.g.*, **Fig. 1B**, **Fig. S2**). Pangenome GWAS also identified a large set of 73/65 accessory genes significantly associated with Molassodon/Typical, distinguishing the ecospecies structure from that of *H. pylori* Hardy (**Fig. 1C**; **Table S4**). These analyses confirmed that Molassodon is an ecospecies, differentiated by only 48/~4800 core genes (at least 5 SNPs with F_{st} ≥0.5 and $-\log_{10}(p) \geq 10$) and over 70 accessory genes (F_{st} ≥0.5 and $-\log_{10}(p) \geq 10$, also requiring >90% or <10% frequency in Molassodon).

Trees generated from 46 differentiated core genes and a set of fully undifferentiated genes (those with no SNPs with F_{st} ≥0.05; **Table S3**) showed different topologies, as we saw previously for *H. pylori* Hardy [3]. The undifferentiated gene tree was star-like and reflected geographic populations [12], with Molassodon strains from three populations dispersed amongst Typical strains (**Fig. 1D,** *top*). Based on differentiated genes, all Molassodon strains instead clustered together on an extended branch, no matter which population they were from (**Fig. 1D,** *bottom*). Therefore, Molassodon genes have a different evolutionary path, despite extensive ongoing genetic exchange between ecospecies in most of the genome.

The Molassodon ecospecies is much older than detectable clonal lineages that share ancestry by direct descent. For example, there are "deep clones" of 21 strains visible in the phylogenetic tree of non-redundant strains, all assigned to the VppAsia population (**Fig. S1A**). The distribution of pairwise differences implies that the strains in the deep clone each share around 7% of their genome by direct descent on average, implying that their common ancestor was likely around 100,000 years ago, based on a recombination clock (**Table S1; Figs. S1D & S1E**) [12]. Molassodon is much older according to the same approach, with the average shared ancestry being less than 1%. There are 232 Molassodon strains in the same dataset including 220 from the VppAsia population and the same recombination clock would imply that their common ancestor existed around a million years ago but the assumptions of the method break down at this time depth. Consistent with its antiquity, Molassodon is found on three of the four different populations in our dataset (VppAsia, VppNAmerica1 and VppNAmerica2) (**Figs. S1A & S1B**).

The pattern of differentiation between Molassodon and Typical strains is qualitatively different from that found for clones or geographic populations. Deep clones and populations are differentiated genome-wide, while Molassodon is undifferentiated from typical strains for more than 90% of the genome, but has very high differentiation for less than 1% of the genome (**Fig. 1A**; **Figs. S3 & S4**). Genome wide differentiation in clones and populations can be attributed to limited genetic exchange, while Molassodon has been freely exchanging genes with Typical strains in most of the genome since its ancient origin and only selected regions of the genome are differentiated.

Further inspection showed that there are three types of Molassodon-differentiated genes based on their mode of evolution (**Figs. 1E & 1F; Fig. S5**). Most core genes are evolving by point mutation with a clear signature of branch-lengthening vs. other *V. parahaemolyticus* (e.g., *lafK*; **Fig. 1E**), 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. 1E**). However, we also identified six core genes with evidence of allelic exchange with other species (*V. antiquarius* and *V. rotiferianus*), including *lafA*, encoding the lateral flagellin, and several genes in the *tonB3* transporter locus (**Fig. 1E; Fig. S5B; Table S3)**.

As we previously noted [10], many of the Molassodon accessory genes (here, 59/73 in encoded by the ecospecies; **Table S5**) 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. Our inspection suggests that these represent continued introgression events from other *Vibrio* species or other marine bacteria (**Table S5**). Although perfect matches were not identified in the database, with the most common closest match (nine genes) being *V. antiquarius*. This locus is the basis of O:K serotyping in the species [13] and despite high predicted serotype diversity in our dataset, we found that the O:K locus of Molassodon is present in only a handful of versions (OL2, OL5, OL11 and KL3, KL28) (**Fig. S6A & S6B; Table S6**).

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 [14–16]. 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

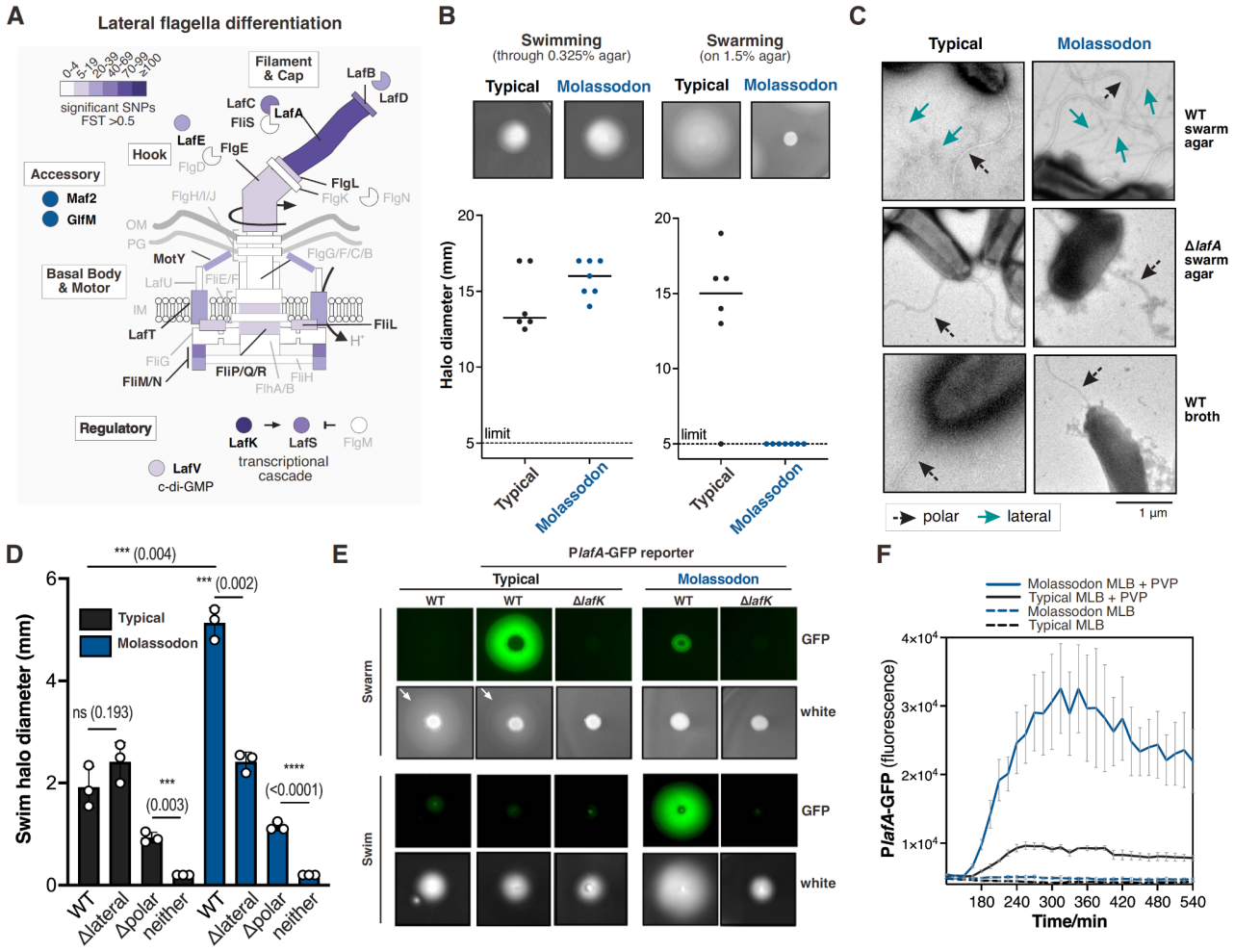

Figure 2. Molassodon uses lateral flagella for swimming in viscous media rather than swarming. (A) Schematic of lateral flagella differentiation in Molassodon strains. Differentiated: F_{st} ≥0.5 and -log10(p) ≥10; Molassodon vs Typical GWAS. **(B)** *Top*: Representative swim/swarm halos for Typical and Molassodon strains. *Bottom*: Halo measurements. Representative of 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**.

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.

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. S7A; Table S8**). 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 (**Fig. S7B**). 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. S7C; Table S9**). 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 [17], we tested the effect of adding PVP (polyvinylpyrrolidone), previously used to explore the effect of viscosity on motility in several *Vibrio* species [6,17,18], 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. S7D**). 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. S8A**). 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 S10**; **Fig. S8B & S8C**). 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. S8C; Table S12**) but not swimming (**Fig. 3A; Table S11**). 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 VI 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. S8B**). 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 [19]. Because motility can increase T6SS efficacy [20–22], 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-

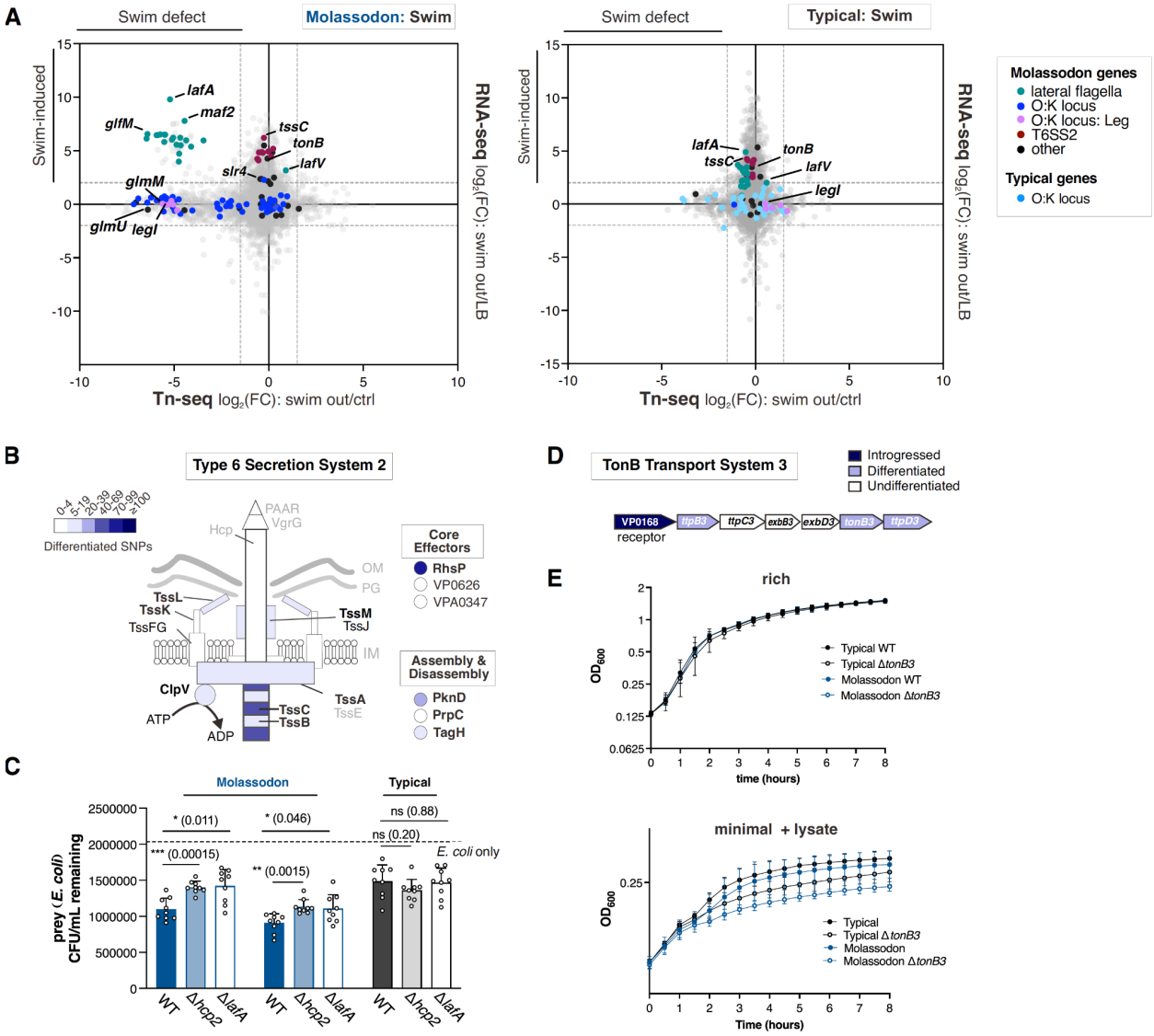

Figure 3. Validation and refinement of the Molassodon laboratory phenotype with Tn- and RNA-seq. (A) Tn-seq and RNA-seq profiles for Molassodon and Typical strains during swimming. Tn-seq: RIMD 2210633 (Typical) and DFS-0010 (Molassodon), $\log_2$FC Swim halo vs. Swim control. RNA-seq: mean of three 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. S8B** and **Tables S10-11**. **(B)** Schematic of T6SS2 with differentiated components indicated. Differentiated: F_{st} ≥0.5 and -log10(p) ≥10); Molassodon vs Typical GWAS. **(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.

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, encode three TonB systems (TonB1-TonB3, according to *V. vulnificus* nomenclature [23]) (**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. S5B**), 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 growth trend for WT strains (**Fig. S9**), 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 VI 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. S10A**) 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. S10A**). 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 and induced during Molassodon swimming (**Fig. 3A & 4A**). Along with upstream *lafA*, the genes show hallmarks of introgression (**Fig. 1D & Fig. S10B**). Based on homology to *Shewanella* genes [24], 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 carries a Leg derivative [25].

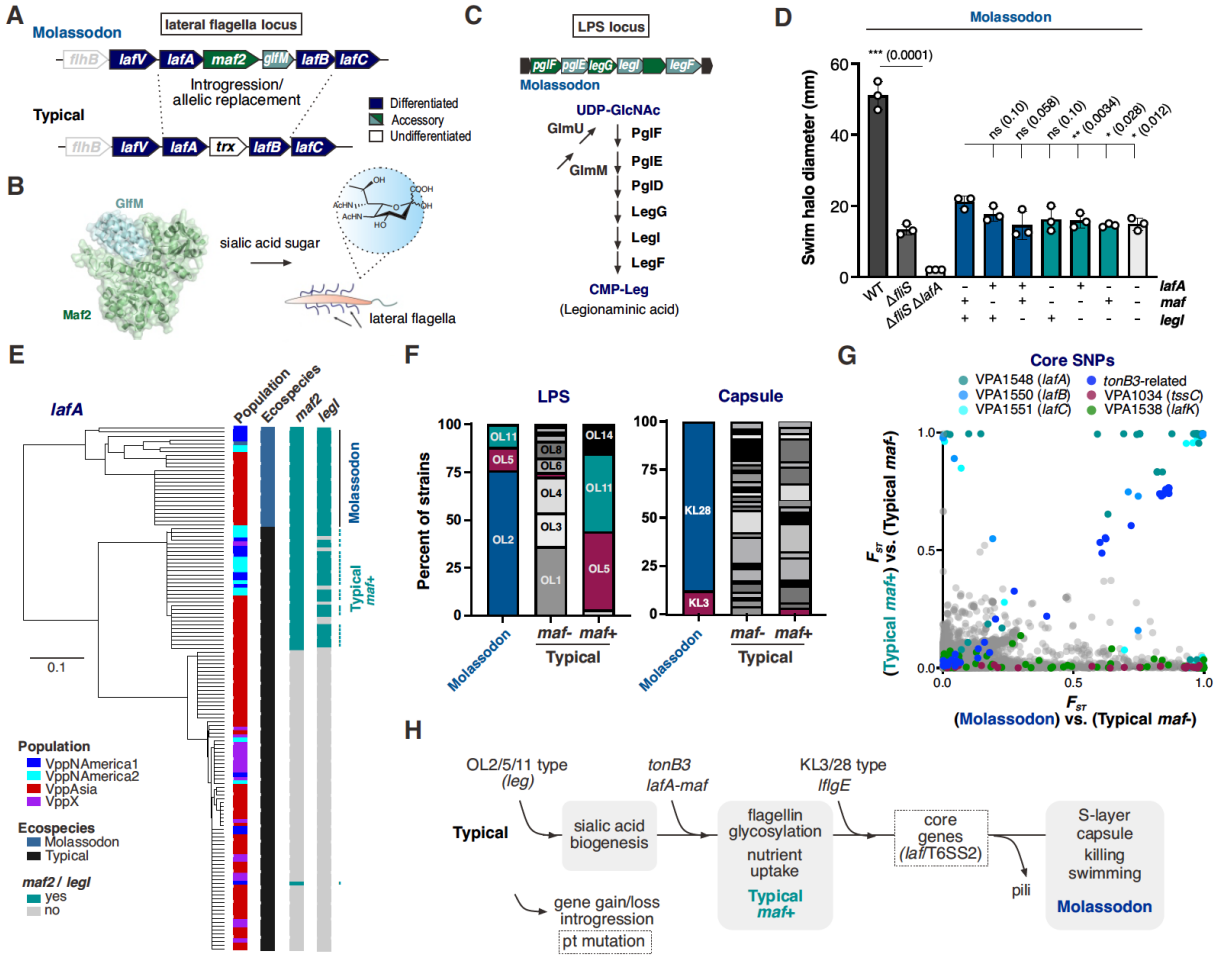

Figure 4. Convergent coadaptation after *maf* introgression. (A) Genomic context of the Molassodon accessory lateral flagella genes *maf2-glfM* and related region in non-introgressed strains. **(B)** Predicted complex formed by Maf2-GlfM and role of the proteins in flagellin glycosylation. The complex was predicted with Alphafold Multimer at Colabfold [26,27] and visualized with Pymol. **(C)** Molassodon genes encoding a putative Leg (legionaminic acid) biosynthesis pathway. The pathway is based on [28]). CMP: cytidine monophosphate. **(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)** Neighbour-joining **p**hylogenetic 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 S6**. **(G)** F_{st} (fixation index) values for (Typical *maf*+) vs. (Typical *maf*-) plotted against corresponding values for (Molassodon) vs. (Typical *maf*-) for core genome SNPs. Key functions identified in this and previous [10,29] work are highlighted. Accessory genes: **Fig. S13A**. **(H)** Evolutionary model for assembly of Molassodon.

Our inferences about the function of two accessory genes encoded downstream of *lafA* in Molassodon strains (**Fig. 1C**) was also confirmed by direct molecular-biological investigation. 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. S10A**). 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.

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 S5**). 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. S13B**). 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 similar to a version of the gene found in *V. antiquarius* than to the Molassodon version (**Fig. S10B**).

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 variants associated with *maf* in Typical strains. 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*- Typical strains also included several in the operon encoding the TonB3 transporter investigated above (**Fig. 3E**). 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. S12A & S12B**) 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 [29], the cellulose biosynthesis cluster is in strong negative LD (linkage disequilibrium) with T6SS1, and accordingly shows an opposite pattern of frequencies (**Fig. S13A**).

Notwithstanding the similarities with Molassodon strains in *maf*, *tonB3*, and the LPS locus, we found that all intermediate strains could swarm (**Fig. S13D**) 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. S13E**), 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, based either on GWAS or F_{st} (**Fig. 4G; Figs. S11 & S13A**). 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. S13D**), 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 serotypes, the capsule locus shows a distinct pattern for Molassodon strains (**Fig. 4F**), consistent with a subsequent remodelling of cell envelope components. The Molassodon capsule locus encodes an ABC transporter-dependent capsule, unlike the Wzx/Wzy-dependent version of Typical strains [30], and perhaps most strikingly, also might encode an abundant S-layer, based on functional annotation and synteny analysis of a putative Type II secretion system and adjacent protein with structural homology to S-layer proteins identified in other marine bacteria (**Fig. S14**) [31]. S-layers are proteinaceous, crystalline layers expressed on the surface of some bacteria and many archaea that can protect from predation/phages and confer unique surface properties [31]. Finally, all Molassodon strains have lost a large island (VPA0713-31; **Fig. 1C**), 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 affected swimming (**Fig. S10A**), 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 identified and refined a laboratory phenotype for the Molassodon ecospecies, validating the ecological linkage of these genes [3,10,29] 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.

First, we defined the Molassodon genotype in detail, showing differentiation in 48 core and 73 (present) accessory genes (**Figs. 1A & 1C**). This genotype is found on multiple otherwise unrelated V. parahaemolyticus backgrounds, including three out of four genetically differentiated oceanic populations, implying that it encodes a robust evolutionary strategy, applicable in many locations. The clean, sharp peaks we obtain in our differentiation scan (**Fig. 1A**) allows us to identify which loci are selected to be differentiated between ecospecies with high sensitivity and specificity. These clean

peaks are facilitated by our large sample of non-redundant Typical and Molassodon genomes as well as extensive recombination between Molassodon and Typical strains over a history of divergence that is at least several hundred thousand years old. Clones or geographical populations have more limited recombination, as well as less clearcut ecological selection and consequently give less distinct peaks (**Figs. S3 & S4**). Our peaks are also narrower than most of the equivalent peaks found in eukaryotic model systems such as sticklebacks, which typically contain many genes [2]. Our sharp genetic peaks increase accuracy in every stage of functional and ecological inference.

We next 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 VI secretion system and the TonB nutrient uptake system. Based on previous findings in other bacteria linking T6SS effectors and TonB-mediated micronutrient uptake [32–34], 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. 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 VI-mediated bacterial antagonism and its ecology [35], but until recently, it has been hard to confirm that killing was an effective foraging strategy [8]. 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 VI secretion systems [8], although mechanisms to actively trap T6SS prey with "grappling hooks" have been reported [36]. Our experimental and coadaptation results provide evidence for flagella-driven predation, whose effectiveness could be enhanced in a viscous environment by limited nutrient diffusion [37]. 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 as well as showing how some steps of this evolution can be reconstructed. For example, we identified two independent imports of neighboring *lafA, maf2 and glfM* genes, one leading to Molassodon and the other to a much simpler coadapted genotype, which represents an earlier stage of functional divergence [38]. We functionally link these imports to additional 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* [25], this provides a feasible pathway for stepwise assembly of part of the complex phenotype (**Fig. 4H**).

The distinct [30] 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 [39]. 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 [6–8,40,41].

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 [42], 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 [43] 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 [42] and may precipitate speciation [10].

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 genomics to identify groups of strains with distinct adaptations, before attempting to infer biological and niche differences based on metadata, experiments, and/or functional annotation [44,45]. We have here used ecospecies to extend the approach from recent to complex adaptation. Powerful molecular genetic approaches available in microbes such as Tn-seq allow experimental linkage of phenotype with genotype, making bacterial ecospecies an enticing new model for functional ecological genomics.

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 (**Fig. S12**) and are detected 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 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) [46], followed by genome assembly with Flye (v2.9-b1768) [47]. 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 [48]. Genomes were annotated using Prokka version 1.14.6 [49] and a pangenome was built using Panaroo version 1.3.3 [50].

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 1,550 public non-redundant *V. parahaemolyticus* genomes (**Table S1**). We used snp-sites [51] to identify single nucleotide polymorphisms (SNPs) from whole-genome alignments. Next, we calculated the SNP distance matrix and performed hierarchical clustering using the complete linkage method. A 10,000 SNP distance threshold 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 [10] to define Molassodon strains (**Fig. S1**). For the *maf* GWAS based on *maf*, we defined strain "phenotype" as "*maf*+" or "*maf*-" based on detection of Molassodon *maf2* by BLAST with Molassodon *maf2*.

For the SNP-based GWAS, we employed snp-sites [51] to extract SNPs from whole-genome alignments, generating a variant call format (VCF) file. Association testing with either "Molassodon" or "*maf*+" was subsequently performed using pyseer [52], which converted inputted multiallelic SNPs into pseudo-bi-allelic categories of major allele or "other", using default parameters. To further assess genetic differentiation between the groups, we calculated fixation index (F_{st}) values. Weir and Cockerham's F_{st} [53] was computed using vcftools [54] on pseudo-haploidized bacterial SNP and accessory gene data, comparing focal strain groups (*e.g.*, Molassodon) against all others. The analysis assumed clonality and negligible within-strain variation.

For the gene-based GWAS (presence/absence), we first used panaroo [50] 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 F_{st} 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 threshold of F_{st}≥0.5 and -log10(p) ≥10. Differentiated core genes had at least 5 SNPs with F_{st}≥0.05 and -log10(p) ≥10, while undifferentiated genes were defined as those with zero SNPs with F_{st}≥0.05 and -log10(p) ≥10. For the gene-based GWAS, the frequency of a gene in Molassodon strains was also required to be at least 90% or less than 10%.

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 [55]. KEGG genomes are listed in **Table S1**.

Phylogenetic analysis. To generate phylogenetic trees, a set of publicly available genomes plus nine newly-sequenced genomes (see above) were used, 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 (*lafA*) 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 using TreeBest. To generate trees based on differentiated/undifferentiated genes, we extracted all SNPs within these regions and constructed an Approximate Maximum Likelihood Tree phylogenetic tree, again using FastTree [56]. Trees were visualized using iTOL [55].

Analysis core genome F_{st} in populations and deep clones. To identify clonal lineages among the 1,550 non-redundant *V. parahaemolyticus* strains, we first calculated a pairwise SNP distance matrix and conducted hierarchical clustering using the single linkage method. Clonal groups were defined by cutting the resulting dendrogram at SNP distance thresholds of 45 kb and 50 kb. The largest putative clonal groups within the VppAsia population obtained under each threshold consisted of 18 (45 kb) and 21 (50 kb) strains, respectively.

To control for sample size effects, we randomly selected 18 and 21 strains from the VppAsia Molassodon lineage to match the clonal groups. For each case, a background population of 950 non-Molassodon VppAsia strains was used to minimize confounding due to broader population structure.

We conducted both genome-wide association studies (GWAS) and FST analyses on the following four pairwise comparisons:

(1) 18-strain clonal group (45 kb threshold) vs. 950 non-Molassodon VppAsia strains
(2) 21-strain clonal group (50 kb threshold) vs. 950 non-Molassodon VppAsia strains
(3) 18 randomly-selected VppAsia Molassodon strains vs. 950 non-Molassodon VppAsia strains
(4) 21 randomly-selected VppAsia Molassodon strains vs. 950 non-Molassodon VppAsia strains

Each pair was treated as a binary phenotype in GWAS and as two distinct populations for F_{st} estimation.

To assess broader patterns of population differentiation and to provide a comparative framework for the Molassodon-based analysis, we constructed two additional datasets for population-level F_{st} estimation:

(1) VppAsia (n = 1,209) vs. VppX (n = 176)

(2) VppAsia (n = 1,209) vs. VppNAmerica (n = 165; combining both VppNAmerica1 and VppNAmerica2)

In each case, F_{st} was calculated between the two groups to quantify genetic differentiation across population lineages.

Coalescence time estimate. We employed a coalescence time estimation approach consistent with the method described previously [29], which combines the estimation of genetic divergence between

strains using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm with the calculation of effective population size (N_e) to derive the coalescence time (T).

By analyzing pairwise distances between strains within different populations of *V. parahaemolyticus*, the expected divergence for each pair of strains, based on their common ancestor, was calculated using the following formula:

$$d_{\text{unrelated}}(1 - \exp(-4 \cdot N_e \cdot r \cdot (\frac{1}{m} - \frac{1}{n})))$$

where $d_{\text{unrelated}}$ represents the distance between unrelated strains, N_e is the effective population size, r is the recombination rate, m is the number of ancestors, and n is the sample size. The observed data is then fitted to this model to estimate $N_e r$, which is subsequently used to calculate the coalescence time T, following the formula:

$$T = \frac{2 \cdot (1 - \frac{1}{n})}{r_{\text{pertime}}} \cdot N_e r$$

To refine our analysis within a single population (VppAsia), we considered strains without any detectable clonal relatedness, which exhibited a median of 54,875 SNPs distinguishing them from the rest of the core genome. Accordingly, we defined $d_{\text{unrelated}}$= 54,875. For the recombination rate r, we adopted a value of 1.7e-4 per site per year, as previously reported [12]. We then focused our analysis on Molassodon strains belonging to the VppAsia lineage, comprising a total of 220 strains.

Coalescent time analyses were subsequently performed using three clonal datasets::
(1) 18-strain clonal group (45 kb SNP-distance threshold)
(2) 21-strain clonal group (50 kb SNP-distance threshold)
(3) 220 VppAsia Molassodon strains

Additional bioinformatic analysis. *In silico* serotyping was performed using the *V. parahaemolyticus* Kaptive database [57,58]. Synteny analysis was performed with webflags (www.webflags.se) [59]. Structures of Slr4 homologues and protein complexes (Maf2 + GlfM) were predicted with AlphaFold2 or Alphafold multimer at Colabfold [26,27]. Structures were visualized with Pymol version 3.0.2. RNA-seq coverage was displayed with Integrated Genome Browser [60].

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 S7**). 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 S7**) 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 [61]. For

swarming, we used either previously described BHI agar (1.5%) [61], 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 S14**. Oligonucleotide primers (Sangon) are listed in **Table S15**. 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 [62], as described previously [61,63]. As an example, the deletion of *lafK* in RIMD 2210633 is described. Oligonucleotides and plasmids used for other mutants are listed in **Tables S14** & **S15**. 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 strains, the two genes with their native RBS were amplified with short, overlapping sequences to the pMMB207 plasmid using DFO-0942/0943 from DFS-0010 gDNA. The pMMB207 plasmid backbone carrying the Ptac promoter was amplified by PCR with DFO-0437/0438. The two l fragments were then assembled into a plasmid by ligation-independent cloning (CPEC, circular polymerase extension cloning) [64] 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 λ*pir* into the appropriate recipient *V. parahaemolyticus* strain.

sfGFP reporter construction. To generate a transcriptional reporter for the Molassodon *lafA* (DFS0010_04645) 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 [65] 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 [66] 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 [67]. 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 [68]. Reads were then mapped to the corresponding reference genome using Segemehl version 0.3.4 with default parameters [69]. FeatureCounts version 2.0.3 was used for gene expression quantification with parameters "-B -C -f -p -s 1 -O --fraction --countReadPairs --ignoreDup –primary" [70]. Differential expression was performed using R package DESeq2 version 1.42.1 using the output from featureCounts without normalization [71]. 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 [72]. Gene set classifications include GO terms [73], KEGG pathways, KEGG KO [74], COG categories [75], 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 [76] and Julia. Plots were generated using Julia package Makie.jl [77].

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[78] 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 [61] 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 [79], 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) [80]. Subsequently, the read counts of each insertion site were tallied using the TraDIS toolkit [81]. To compare the changes in the relative abundance of insertions in each 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.

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 1x 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 S7**). 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 30°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 OD_{600} of 0.5. Each strain was grown in triplicate in 200 µl of MLB or M9 (salts only + NaCl to 2%; 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 for experimental data were performed using Graphpad prism.

Acknowledgements

We thank Qiyao Wang, Yiquan Zhang, Kim Orth, and Melanie Blokesch for providing strains and advice, and Elise Tourrette for assistance with GWAS. **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 (#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. Foote, A.D., Vijay, N., Ávila-Arcos, M.C., Baird, R.W., Durban, J.W., Fumagalli, M., Gibbs, R.A., Hanson, M.B., Korneliussen, T.S., Martin, M.D., et al. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nat. Commun. *7*, 11693. 10.1038/ncomms11693.
2. Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature *484*, 55–61. 10.1038/nature10944.
3. Tourrette, E., Torres, R.C., Svensson, S.L., Matsumoto, T., Miftahussurur, M., Fauzia, K.A., Alfaray, R.I., Vilaichone, R.-K., Tuan, V.P., *Helicobacter* Genomics Consortium, et al. (2024). An ancient ecospecies of *Helicobacter pylori.* Nature. 10.1038/s41586-024-07991-z.
4. Thorell, K., Muñoz-Ramírez, Z.Y., Wang, D., Sandoval-Motta, S., Boscolo Agostini, R., Ghirotto, S., Torres, R.C., HpGP Research Network, Falush, D., Camargo, M.C., et al. (2023). The Helicobacter pylori Genome Project: insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. Nat. Commun. *14*, 8184. 10.1038/s41467-023-43562-y.
5. Broberg, C.A., Calder, T.J., and Orth, K. (2011). *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. Microbes Infect. *13*, 992–1001. 10.1016/j.micinf.2011.06.013.
6. Grognot, M., Nam, J.W., Elson, L.E., and Taute, K.M. (2023). Physiological adaptation in flagellar architecture improves *Vibrio alginolyticus* chemotaxis in complex environments. Proc Natl Acad Sci USA *120*, e2301873120. 10.1073/pnas.2301873120.
7. Speare, L., Cecere, A.G., Guckes, K.R., Smith, S., Wollenberg, M.S., Mandel, M.J., Miyashiro, T., and Septer, A.N. (2018). Bacterial symbionts use a Type VI secretion system to eliminate competitors in their natural host. Proc Natl Acad Sci USA *115*, E8528–E8537. 10.1073/pnas.1808302115.
8. Stubbusch, A.K.M., Peaudecerf, F.J., Lee, K.S., Paoli, L., Schwartzman, J., Stocker, R., Basler, M., Schubert, O.T., Ackermann, M., Magnabosco, C., et al. (2025). Antagonism as a foraging strategy in microbial communities. Science *388*, 1214–1217. 10.1126/science.adr8286.
9. Freitas, C., Glatter, T., and Ringgaard, S. (2020). The release of a distinct cell type from swarm colonies facilitates dissemination of *Vibrio parahaemolyticus* in the environment. ISME J. *14*, 230–244. 10.1038/s41396-019-0521-x.
10. Cui, Y., Yang, C., Qiu, H., Wang, H., Yang, R., and Falush, D. (2020). The landscape of coadaptation in *Vibrio parahaemolyticus*. eLife *9*. 10.7554/eLife.54136.
11. Homma, M., Nishikino, T., and Kojima, S. (2022). Achievements in bacterial flagellar research with focus on *Vibrio* species. Microbiol. Immunol. *66*, 75–95. 10.1111/1348-0421.12954.
12. Yang, C., Pei, X., Wu, Y., Yan, L., Yan, Y., Song, Y., Coyle, N.M., Martinez-Urtaza, J., Quince, C., Hu, Q., et al. (2019). Recent mixing of *Vibrio parahaemolyticus* populations. ISME J. *13*, 2578–2588. 10.1038/s41396-019-0461-5.
13. Chen, Y., Dai, J., Morris, J.G., and Johnson, J.A. (2010). Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. BMC Microbiol. *10*, 274. 10.1186/1471-2180-10-274.
14. McCarter, L., Hilmen, M., and Silverman, M. (1988). Flagellar dynamometer controls swarmer cell differentiation of *V. parahaemolyticus*. Cell *54*, 345–351. 10.1016/0092-8674(88)90197-3.

15. McCarter, L., and Silverman, M. (1990). Surface-induced swarmer cell differentiation of *Vibrio parahaemolyticus*. Mol. Microbiol. *4*, 1057–1062. 10.1111/j.1365-2958.1990.tb00678.x.

16. Gode-Potratz, C.J., Kustusch, R.J., Breheny, P.J., Weiss, D.S., and McCarter, L.L. (2011). Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. Mol. Microbiol. *79*, 240–263. 10.1111/j.1365-2958.2010.07445.x.

17. Kawagishi, I., Imagawa, M., Imae, Y., McCarter, L., and Homma, M. (1996). The sodium-driven polar flagellar motor of marine *Vibrio* as the mechanosensor that regulates lateral flagellar expression. Mol. Microbiol. *20*, 693–699. 10.1111/j.1365-2958.1996.tb02509.x.

18. Septer, A.N., Sharpe, G., and Shook, E.A. (2023). The *Vibrio fischeri* type VI secretion system incurs a fitness cost under host-like conditions. BioRxiv. 10.1101/2023.03.07.529561.

19. Tang, L., Dong, S., Rasheed, N., Wu, H.W., Zhou, N., Li, H., Wang, M., Zheng, J., He, J., and Chao, W.C.H. (2022). *Vibrio parahaemolyticus* prey targeting requires autoproteolysis-triggered dimerization of the Type VI secretion system effector RhsP. Cell Rep. *41*, 111732. 10.1016/j.celrep.2022.111732.

20. Booth, S.C., Meacock, O.J., and Foster, K.R. (2024). Cell motility empowers bacterial contact weapons. ISME J. *18*. 10.1093/ismejo/wrae141.

21. Gallegos-Monterrosa, R., and Coulthurst, S.J. (2021). The ecological impact of a bacterial weapon: microbial interactions and the Type VI secretion system. FEMS Microbiol. Rev. *45*. 10.1093/femsre/fuab033.

22. Ho, B.T., Dong, T.G., and Mekalanos, J.J. (2014). A view to a kill: the bacterial Type VI secretion system. Cell Host Microbe *15*, 9–21. 10.1016/j.chom.2013.11.008.

23. Kustusch, R.J., Kuehl, C.J., and Crosa, J.H. (2012). 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. 10.1128/JB.00155-12.

24. Unay, J., Kint, N., and Viollier, P.H. (2024). 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. 10.1016/j.cub.2024.05.058.

25. Hashii, N., Isshiki, Y., Iguchi, T., and Kondo, S. (2003). Structural analysis of the carbohydrate backbone of *Vibrio parahaemolyticus* O2 lipopolysaccharides. Carbohydr. Res. *338*, 1063–1071. 10.1016/S0008-6215(03)00078-8.

26. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. 10.1038/s41592-022-01488-1.

27. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A.W., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. BioRxiv. 10.1101/2021.10.04.463034.

28. Morrison, M.J., and Imperiali, B. (2014). The renaissance of bacillosamine and its derivatives: pathway characterization and implications in pathogenicity. Biochemistry *53*, 624–638. 10.1021/bi401546r.

29. Cui, Y., Yang, X., Didelot, X., Guo, C., Li, D., Yan, Y., Zhang, Y., Yuan, Y., Yang, H., Wang, J., et al. (2015). Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen *Vibrio parahaemolyticus*. Mol. Biol. Evol. *32*, 1396–1410. 10.1093/molbev/msv009.

30. Bian, S., Zeng, W., Li, Q., Li, Y., Wong, N.-K., Jiang, M., Zuo, L., Hu, Q., and Li, L. (2020). Genetic Structure,

Function, and Evolution of Capsule Biosynthesis Loci in *Vibrio parahaemolyticus*. Front. Microbiol. *11*, 546150. 10.3389/fmicb.2020.546150.

31. Ali, S., Jenkins, B., Cheng, J., Lobb, B., Wei, X., Egan, S., Charles, T.C., McConkey, B.J., Austin, J., and Doxey, A.C. (2020). Slr4, a newly identified S-layer protein from marine Gammaproteobacteria, is a major biofilm matrix component. Mol. Microbiol. *114*, 979–990. 10.1111/mmi.14588.
32. Si, M., Zhao, C., Burkinshaw, B., Zhang, B., Wei, D., Wang, Y., Dong, T.G., and Shen, X. (2017). Manganese scavenging and oxidative stress response mediated by Type VI secretion system in Burkholderia thailandensis. Proc Natl Acad Sci USA *114*, E2233–E2242. 10.1073/pnas.1614902114.
33. Wang, T., Si, M., Song, Y., Zhu, W., Gao, F., Wang, Y., Zhang, L., Zhang, W., Wei, G., Luo, Z.-Q., et al. (2015). Type VI secretion system transports Zn^{2+} to combat multiple stresses and host immunity. PLoS Pathog. *11*, e1005020. 10.1371/journal.ppat.1005020.
34. Si, M., Wang, Y., Zhang, B., Zhao, C., Kang, Y., Bai, H., Wei, D., Zhu, L., Zhang, L., Dong, T.G., et al. (2017). The Type VI Secretion System Engages a Redox-Regulated Dual-Functional Heme Transporter for Zinc Acquisition. Cell Rep. *20*, 949–959. 10.1016/j.celrep.2017.06.081.
35. Granato, E.T., Meiller-Legrand, T.A., and Foster, K.R. (2019). The evolution and ecology of bacterial warfare. Curr. Biol. *29*, R521–R537. 10.1016/j.cub.2019.04.024.
36. Lien, Y.-W., Amendola, D., Lee, K.S., Bartlau, N., Xu, J., Furusawa, G., Polz, M.F., Stocker, R., Weiss, G.L., and Pilhofer, M. (2024). Mechanism of bacterial predation via ixotrophy. Science *386*, eadp0614. 10.1126/science.adp0614.
37. Nadell, C.D., and Marx, C.J. (2025). Bacteria poison and eat their neighbors. Science *388*, 1134–1135. 10.1126/science.ady7008.
38. Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., Polz, M.F., and Alm, E.J. (2012). Population genomics of early events in the ecological differentiation of bacteria. Science *336*, 48–51. 10.1126/science.1218198.
39. Hehemann, J.-H., Arevalo, P., Datta, M.S., Yu, X., Corzett, C.H., Henschel, A., Preheim, S.P., Timberlake, S., Alm, E.J., and Polz, M.F. (2016). Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. Nat. Commun. *7*, 12860. 10.1038/ncomms12860.
40. Zhao, W., Caro, F., Robins, W., and Mekalanos, J.J. (2018). Antagonism toward the intestinal microbiota and its effect on *Vibrio cholerae* virulence. Science *359*, 210–213. 10.1126/science.aap8775.
41. Otto, S.B., Servajean, R., Lemopoulos, A., Bitbol, A.-F., and Blokesch, M. (2024). Interactions between pili affect the outcome of bacterial competition driven by the type VI secretion system. Curr. Biol. *34*, 2403-2417.e9. 10.1016/j.cub.2024.04.041.
42. Yang, C., Huang, H., Wang, N., Didelot, X., Yang, R., Cui, Y., and Falush, D. (2025). Macrogenetic atlas of prokaryotes. BioRxiv. 10.1101/2025.02.26.640246.
43. Peoples, N., Burns, M.D., Mihalitsis, M., and Wainwright, P.C. (2025). Evolutionary lability of a key innovation spurs rapid diversification. Nature. 10.1038/s41586-025-08612-z.
44. Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J., and Polz, M.F. (2019). A reverse ecology approach based on a biological definition of microbial populations. Cell *178*, 820-834.e14. 10.1016/j.cell.2019.06.033.
45. Shapiro, B.J., and Polz, M.F. (2014). Ordering microbial diversity into ecologically and genetically

cohesive units. Trends Microbiol. *22*, 235–247. 10.1016/j.tim.2014.02.006.

46. GitHub - rrwick/Filtlong: quality filtering tool for long reads https://github.com/rrwick/Filtlong.
47. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546. 10.1038/s41587-019-0072-8.
48. Wick, R.R., Judd, L.M., Cerdeira, L.T., Hawkey, J., Méric, G., Vezina, B., Wyres, K.L., and Holt, K.E. (2021). Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol. *22*, 266. 10.1186/s13059-021-02483-z.
49. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. 10.1093/bioinformatics/btu153.
50. Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R.A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. *21*, 180. 10.1186/s13059-020-02090-4.
51. Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., and Harris, S.R. (2016). SNP-sites : rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genom. *2*, e000056. 10.1099/mgen.0.000056.
52. Lees, J.A., Galardini, M., Bentley, S.D., Weiser, J.N., and Corander, J. (2018). pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics *34*, 4310–4312. 10.1093/bioinformatics/bty539.
53. Weir, B.S., and Cockerham, C.C. (1984). Estimating f-statistics for the analysis of population structure. Evolution *38*, 1358–1370. 10.1111/j.1558-5646.1984.tb05657.x.
54. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. 10.1093/bioinformatics/btr330.
55. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. *52*, W78–W82. 10.1093/nar/gkae268.
56. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. *26*, 1641–1650. 10.1093/molbev/msp077.
57. Wick, R.R., Heinz, E., Holt, K.E., and Wyres, K.L. (2018). Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. J. Clin. Microbiol. *56*. 10.1128/JCM.00197-18.
58. van der Graaf-van Bloois, L., Chen, H., Wagenaar, J.A., and Zomer, A.L. (2023). Development of Kaptive databases for *Vibrio parahaemolyticus* O- and K-antigen genotyping. Microb. Genom. *9*. 10.1099/mgen.0.001007.
59. Saha, C.K., Sanches Pires, R., Brolin, H., Delannoy, M., and Atkinson, G.C. (2021). FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. Bioinformatics *37*, 1312–1314. 10.1093/bioinformatics/btaa788.
60. Freese, N.H., Norris, D.C., and Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. Bioinformatics *32*, 2089–2095. 10.1093/bioinformatics/btw069.
61. Chimalapati, S., Lafrance, A.E., Chen, L., and Orth, K. (2020). Vibrio parahaemolyticus: Basic Techniques for Growth, Genetic Manipulation, and Analysis of Virulence Factors. Curr. Protoc. Microbiol. *59*, e131.

10.1002/cpmc.131.

62. Mou, X., Spinard, E.J., Driscoll, M.V., Zhao, W., and Nelson, D.R. (2013). H-NS is a negative regulator of the two hemolysin/cytotoxin gene clusters in *Vibrio anguillarum*. Infect. Immun. *81*, 3566–3576. 10.1128/IAI.00506-13.

63. Yang, C., Qiu, H., Svensson, S.L., Ni, C., Gao, S., Jia, Z., Wen, H., Xie, L., Xu, W., Qin, Y., et al. (2025). Wave succession in the pandemic clone of *Vibrio parahaemolyticus* driven by gene loss. Nat. Ecol. Evol. 10.1038/s41559-025-02827-z.

64. Quan, J., and Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. PLoS ONE *4*, e6441. 10.1371/journal.pone.0006441.

65. Corcoran, C.P., Podkaminski, D., Papenfort, K., Urban, J.H., Hinton, J.C.D., and Vogel, J. (2012). Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. Mol. Microbiol. *84*, 428–445. 10.1111/j.1365-2958.2012.08031.x.

66. Morales, V.M., Bäckman, A., and Bagdasarian, M. (1991). A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. Gene *97*, 39–47. 10.1016/0378-1119(91)90007-X.

67. Sharma, C.M., Darfeuille, F., Plantinga, T.H., and Vogel, J. (2007). 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. 10.1101/gad.447207.

68. Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta *2*, e107. 10.1002/imt2.107.

69. Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L.M., Teupser, D., Hackermüller, J., et al. (2014). A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol. *15*, R34. 10.1186/gb-2014-15-2-r34.

70. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. 10.1093/bioinformatics/btt656.

71. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1186/s13059-014-0550-8.

72. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) *2*, 100141. 10.1016/j.xinn.2021.100141.

73. Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The Gene Ontology knowledgebase in 2023. Genetics *224*, iyad031. 10.1093/genetics/iyad031.

74. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci. *28*, 1947–1951. 10.1002/pro.3715.

75. Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., and Koonin, E.V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. *49*, D274–D281. 10.1093/nar/gkaa1018.

76. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*.

10.1093/gigascience/giab008.

77. Danisch, S., and Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia. JOSS *6*, 3349. 10.21105/joss.03349.

78. Wang, Z., Zhu, S., Li, C., Lyu, L., Yu, J., Wang, D., Xu, Z., Ni, J., Gao, B., Lu, J., et al. (2022). Gene essentiality profiling reveals a novel determinant of stresses preventing protein aggregation in *Salmonella*. Emerg. Microbes Infect. *11*, 1554–1571. 10.1080/22221751.2022.2081618.

79. Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe *6*, 279–289. 10.1016/j.chom.2009.08.003.

80. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. 10.1186/gb-2009-10-3-r25.

81. Barquist, L., Mayho, M., Cummins, C., Cain, A.K., Boinett, C.J., Page, A.J., Langridge, G.C., Quail, M.A., Keane, J.A., and Parkhill, J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics *32*, 1109–1111. 10.1093/bioinformatics/btw022.