Functional coherence of bacterial 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 Materia Medica, 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

Organizing bacterial diversity into biologically meaningful units is key to understanding the molecular and evolutionary basis of adaptation. We have recently identified bacterial "ecospecies": strains that share a well-mixed gene pool with the rest of the species in most of the genome, but carry nearly fixed differences in ~100 coadapted genes. We hypothesized this structure reflects ecological selection, but ecospecies have so far been characterized only genomically. Here, using "Molassodon" of the marine bacterium *Vibrio parahaemolyticus,* we provide first experimental evidence linking a multilocus ecospecies genotype to a laboratory phenotype. We find that Molassodon swimming in viscous environments is enhanced by differentiated lateral flagella that no longer drive surface swarming. Using Tn-seq, RNA-seq, and molecular genetics, we link differentiated Type VI secretion and nutrient uptake genes to the swimming phenotype, potentially invoking a natural adaptation related to "hunting". We also identify a convergent genotype also linking lateral flagella and nutrient uptake, suggesting ecological differentiation in *V. parahaemolyticus* occurs along recurring phenotypic axes and involves borrowing traits from a broader *Vibrio* gene pool. Together, our results provide first experimental evidence that ecospecies are functionally coherent units and illustrate how they can provide a window into the complex adaptations that mediate bacterial ecology.

Introduction

The most intensively studied bacteria are human pathogens, and thousands of articles are published each year on their molecular epidemiology, identifying disease causing lineages and characterizing their distribution and pathogenic potential. These studies demonstrate that bacterial phenotypes can be inherited stably by particular lineages for tens or thousands of years [12]. For example in *Vibrio parahaemolyticus*, many isolates that cause human disease belong to a single clonal lineage whose common ancestor is less than 100 years old [13,14], and accordingly has much lower genetic diversity than the species as a whole. Although the ecological role of some pathogen phenotypes, such as the tendency of particular *Neisseria meningitidis* lineages to cause invasive disease, is unknown, many pathogenicity related phenotypes clearly aid bacterial transmission [15,16] and hence allow the clade that carries them to occupy a distinct ecological niche, helping to explain some of their success.

Although molecular epidemiology has been highly successful in characterizing the genetics of pathogens, the biological and evolutionary mechanisms driving bacterial differentiation remain poorly understood, limiting our understanding of how pathogenesis or other traits emerge or of the overall niche structure within species. [17,18]). One theoretical perspective is the stable ecotype model,

which suggests that species are divided into "ecotypes". An ecotype is a clade which is ecologically differentiated from the rest of the species. According to the model, genetic diversity within each ecotype is periodically purged by selection or drift [19]. Ecotypes have also been defined in a more operational way, based on identifying phylogenetic clusters that are sampled preferentially from particular habitat types, implying that they are likely to be associated with a specific niche [19,20].

An alternative to the dominant phylogenetic perspective on ecological differentiation takes a genes-eye view, which can be appropriate if recombination rates are high. Accordingly, Shapiro and Polz (2014) defined ecologically coherent groups based on small genomic regions of differentiation that underlie adaptation [17]. This model was empirically supported by detection of a small number of "ecoSNPs" differentiating *Vibrio* populations living on different size particles [21]. Other approaches first identify ecologically cohesive populations, independent of genetic signatures [22].

Bacterial "ecospecies" represent a newly recognized genetic structure that potentially represents a powerful context to study the molecular basis of ecological adaptation. In a species with two ecospecies, most of the genome is freely recombining, while a small number of regions show nearly fixed differences [3]. This genetic structure shares features with ecotypes and ecoSNPs but uniquely can be recognized without any form of ecological information, including about sampling locations. Nevertheless, identifying a bacterial ecospecies begs a question, namely what is the force maintaining divergence in the differentiated regions in the face of high recombination rates. The name ecospecies reflects what seems to be the only plausible answer, namely that the differentiation is maintained by natural selection. More specifically, we have hypothesized that the differentiated genes are coadapted to each other, which is to say they act together to allow the bacteria to colonize a distinct ecological niche [3].

The ecological relevance of the "Hardy" ecospecies of the gastric pathogen *Helicobacter pylori* is supported by functional annotation of differentiated genes and the isolation sources of Hardy strains. Hardy differs from Ubiquitous *H. pylori* at only ~100~1,500 core genes and a small number of accessory genes [3]. These differences have been maintained for at least 300,000 years across diverse *H. pylori* ancestries, despite the extensive recombination during mixed infections that rapidly erases clonal structure [3,4]. Based on the isolation of Hardy from indigenous human populations and certain animals (e.g., big cats) with meat-biased diets, together with a Hardy-specific urease that uses an iron cofactor that is also found in Helicobacter species that colonize other carnivorous mammals, we proposed that the Hardy genotype represents an adaptation to carnivorous human hosts. However, this hypothesis requires experimental validation, for example by identification of an adaptive laboratory phenotype that encompasses differentiated genes.

Several *Vibrio* species have served as models for studying the genetic basis of ecological differentiation, as well as ecologically-relevant traits such motility, symbiosis, and Type VI secretion-mediated bacterial antagonism [6–9,20–22]. *V. parahaemolyticus* is widespread in coastal and estuarine environments, existing free-living or associated with animals, plankton, or particles, and its range is expanding, likely due to global warming [23]. Unlike most *Vibrio* spp., it encodes a secondary set of lateral flagella linked to surface motility [24]. Some lineages cause seafood-associated gastroenteritis or infect aquaculture species, and accordingly, diversity in the species has been studied primarily

through an epidemiological lens, for example identification of a pandemic clone lineage and other infection-related sequence types [14,25,26].

Our own broader population genomic analyses of *V. parahaemolyticus* have revealed extensive recombination, largely separate geographic gene pools, and evidence for recent population mixing [27–29]. This work suggested that *V. parahaemolyticus* genetic structure is shaped not only by phylogeny or geography but also by ecological differentiation. Exploiting its high recombination rate and large *Ne*, we used genome-wide epistasis scans (GWES) to detect signals of non-randomly associated, coadapted loci, which we termed eco-LD: linkage disequilibrium driven by ecological selection rather than physical linkage [30]. Hierarchical clustering based on these interactions identified EG1a, a distinct, globally distributed group differentiated from other *V. parahaemolyticus* at ~100 genes, and most strikingly those encoding lateral flagella. As GWES is phenotype-independent, how the identified differentiation might impact behaviour or ecology was unclear.

Here, we demonstrate that the EG1a multilocus genotype represents a genetically and functionally coherent ecospecies of *V. parahaemolyticus,* which we now term Molassodon. Molassodon is differentiated from "Typical" *V. parahaemolyticus* by only 48 core and 73 accessory genes, many of which may have been acquired via introgression. We show that Molassodon is old, stable and genetically distinct from clonal, geographic, or phylogenetically defined ecotype structures.

Using the list of Molassodon-differentiated genes, we identify a distinct, coherent laboratory phenotype: unlike Typical *V. parahaemolyticus*, Molassodon induces its differentiated lateral flagella in viscous media, but does not seem to swarm. We leveraged this tractable phenotypic difference, together with RNA-seq and Tn-seq (transposon sequencing), to link this swimming behaviour to differentiated traits beyond lateral flagella, including nutrient uptake and Type VI secretion, suggesting the adaptive advantage encoded by Molassodon loci could be related to active foraging ("hunting") of bacterial prey in viscous environments. Finally, we identify a convergent, intermediate genotype that provides additional support for the functional link between motility and nutrient acquisition in *V. parahaemolyticus* ecology. These genomic and experimental findings reinforce the ecospecies concept as a meaningful organizing principle, and open new avenues for using ecospecies to explore the genetic basis of complex adaptations and the ecological factors driving them in bacteria.

Results

The multilocus genotype identified previously by epistasis analysis is an ecospecies, Molassodon

We previously identified a multilocus genotype in *Vibrio parahaemolyticus* (EG1a) using a genome-wide epistasis scan (GWES) for non randomly-associated, coadapted loci [10]. This involved pairwise Fisher's exact tests between variants in a non-redundant genome dataset from the VppAsia population. This identified a large network of coadapted genes, and hierarchical clustering revealed several strain clusters, of which EG1a was the most distinct. We hypothesized that the variants

underlying EG1a constituted an independently evolving gene pool, and thus an ecospecies, as we had demonstrated for the *H. pylori* Hardy [3]. Based on additional genomics analysis described below, we now designate EG1a as the ecospecies "Molassodon" (based on experiments also described below) and refer to all other strains as "Typical".

The ecospecies genetic signature we observed for *H. pylori* Hardy was small islands of near-fixed differentiation within a freely mixing genome [3]. The signature was detected in several geographic populations, and differentiated genes showed a distinct evolutionary trajectory from the rest of the genome. To confirm that EG1a is an ecospecies, we performed a similar core-genome GWAS (genome-wide association study), comparing Molassodon and Typical strains using a larger non-redundant genome dataset spanning all four populations (VppAsia, VppX, VppNAmerica1, VppNAmerica2 [27,30]. Strains were independently classified as Molassodon/Typical by GWES essentially as performed previously (**Fig. S1A** and **Methods**) [10]. Only ~15 core genome regions showed near-fixed differences between Molassodon and Typical *V. parahaemolyticus* (F_{ST} > 0.9; **Fig. 1A**), corresponding to 46 differentiated core genes (≥5 significant SNPs; $-\log_{10} > 10$; **Tables S1–S4**). Differentiation decayed to near zero within a few hundred base pairs of these loci (**Fig. 1B**; **Fig. S2**), as observed previously for *H. pylori* Hardy [3]. Thus, Molassodon differs from Typical strains at only a small number of loci, with little differentiation elsewhere in the genome, consistent with ecospecies structure.

Phylogenies based on *H. pylori* ecospecies-differentiated core genes were markedly distinct from those based on only undifferentiated genes: the former clustered Hardy strains together on a single long branch, whereas the latter tree showed them dispersed among Ubiquitous strains according to their geographic population [3]. We found that this pattern is recapitulated for *V. parahaemolyticus* Molassodon. A phylogeny based on undifferentiated genes (no SNPs with $F_{ST} \geq 0.05$; n = 3,708; **Table S3**) was star-like (**Fig. 1C**, *top*), as expected for highly recombining *V. parahaemolyticus*, and largely reflected geographic population structure [27]. Molassodon strains, which were detected in three of four geographic populations, were distributed among Typical strains, and according to their assigned geographic population. This indicates that Molassodon is largely undifferentiated from Typical strains outside of coadapted loci. In contrast, a phylogeny based on the 46 ecospecies-differentiated core genes clustered all Molassodon strains together on a single long branch, independent of their population assignment (**Fig. 1C**, *bottom*). Thus, as we observed for *H. pylori*, the small handful of differentiated ecospecies genes have a distinct evolutionary trajectory from the rest of the genome, despite extensive within-species gene flow. Based on this, we concluded that the multilocus EG1a genotype identified by GWES represents a distinct ecospecies, which we now term Molassodon based on functional characterization.

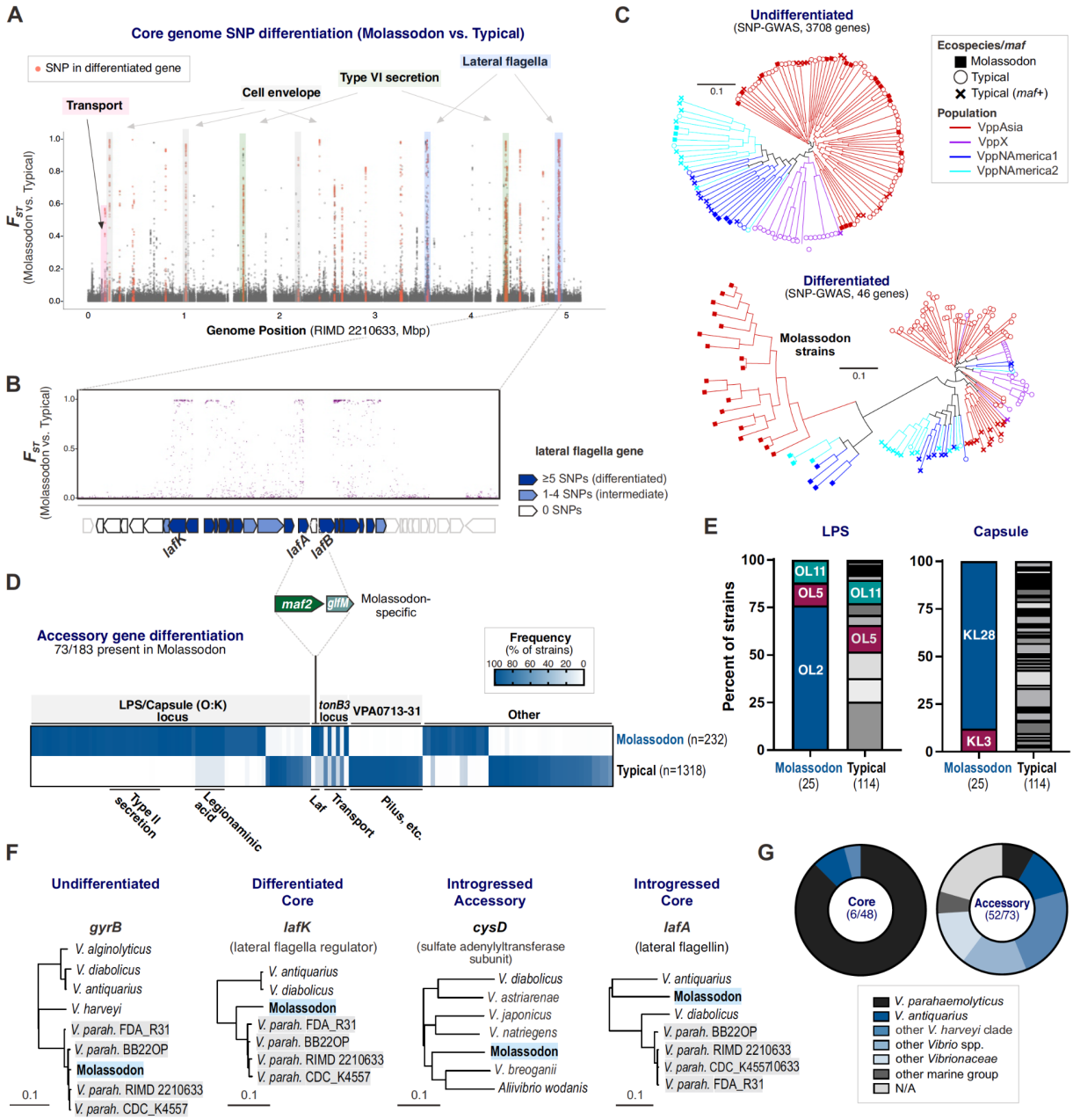

Figure 1. The Molassodon ecospecies genotype. (A) Manhattan plot for core genome SNP (single-nucleotide polymorphism) GWAS of Molassodon (n=223) vs. Typical (n=1318) strains. Significant SNPs: F_{st} (fixation index) ≥0.5 and -log10(p) ≥10. Only those in differentiated genes (5 or more SNPs) are labeled 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)** Approximate Maximum Likelihood 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. **(D)** Frequency of genes identified as significantly differentially

distributed between Molassodon and Typical strains via pangenome GWAS. See also **Table S5**. Significant genes: $F_{ST} \geq 0.5$ and $-\log_{10}(p) \geq 10$, also requiring >90% or <10% frequency in Molassodon. Genes are ordered by putative function. **(E)** Distribution of predicted serotypes for the reduced set of 136 genomes (**Fig. 1C**). Serotypes present in Molassodon strains are coloured. Serotypes were predicted from genome sequences with Kaptive [31,32]. See also **Table S5**. **(F)** 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**. **(G)** Taxon of best KEGG ortholog (by % amino acid identity) for Molassodon-present accessory and core genes. N/A: no match above 40%.

The Molassodon ecospecies is much older than identifiable clonal lineages

In addition to the recent pandemic clonal lineage that causes the large majority of human infections, our global *V. parahaemolyticus* dataset has sets of strains that are more distantly related, but still show signs of common descent. For example we detected a "deep clone" of 21 strains visible in the phylogenetic tree of non-redundant strains used for GWAS, all assigned to the VppAsia population (**Fig. S1A**). The distribution of pairwise differences between strains of the clone implies that they each share on average around 7% of their genome by direct descent, implying that their common ancestor was likely around 100,000 years ago, based on a recombination clock (**Table S1; Figs. S1D & S1E,** see **Methods** for details) [27].

Molassodon is much older according to the same approach, with the average shared ancestry being less than 1%. Using the 220 non-redundantMolassodon strains from the same VppAsia population as the above clone and the same recombination clock, we calculated that their common ancestor existed around a million years ago - although the assumptions of the method break down at this depth of time. Consistent with its antiquity, Molassodon is so far found on three of the four different populations in our dataset (VppAsia, VppNAmerica1 and VppNAmerica2) (**Figs. S1A & S1B**). Furthermore, consistent with the homogenizing effect of recombination over long time periods, Molassodon is undifferentiated from Typical strains for more than 90% of the genome and shows strong differentiation for less than 1% of the genome (**Fig. 1A**; **Figs. S3 & S4**).

Accessory and introgressed traits also distinguish Molassodon and Typical strains

The *H. pylori* Hardy ecospecies genotype includes both core-gene alleles as well as accessory genes (*e.g.*, an iron-cofactored urease) [3]. We therefore expanded on the 46 differentiated core genes by performing a GWAS for accessory gene presence/absence, identifying 73 genes enriched and 65 depleted in Molassodon strains vs. Typical (**Fig. 1D**; **Table S4**). Two of the identified accessory genes are actually core genes in the *tonB3* locus (VP0162/0163) that are sufficiently divergent to be called as "accessory" genes (**Fig. S6**); thus we added these to our core genome list to make a final count of 48.

As noted previously [10], most Molassodon-associated accessory genes (59/73; **Table S5**) are concentrated in a large genomic block. We now note that this block corresponds to the LPS/capsule locus, which underlies O:K serotyping in *Vibrio* spp. [33] (**Fig. S7**; **Table S6**). Despite generally high

serotype diversity in the analyzed dataset for Typical strains, Molassodon strains carry very few O or K types (OL2/5/11; KL3/28) (**Fig. 1E**). Previous work has shown that KL3/28 types are highly divergent, and may be the result of two insertion events from other species [34]. Our inspection likewise suggests that several genes in this locus derive from several introgression events, with closest matches predominantly in other Vibrionaceae, most frequently *V. antiquarius* (**Table S5**).

We identified three evolutionary patterns among Molassodon-differentiated genes (**Fig. 1F,G; Fig. S5**). First, most core genes evolved via point mutation, with the Molassodon version branching together with other *V. parahaemolyticus* strains, albeit often with a slightly extended branch length (*e.g.*, *lafK*). Second, most Molassodon accessory genes (52/73) bear signatures of introgression from other species, as they have closest homologs outside of *V. parahaemolyticus* (*e.g.*, *cysD*). Finally, we also identified six core genes with evidence of allelic exchange with other species (*V. antiquarius* and *V. rotiferianus*), including *lafA* and genes in the *tonB3* locus (**Figs. 1F**, **S5B**; **Table S3**). In sum, this implies that a large proportion of the multigene Molassodon genotype has been acquired from outside of *V. parahaemolyticus*.

Overall, the above systematic analyses 1) confirm that the multilocus genotype EG1a is an ecospecies, which is distinct from a population or a clone, 2) that Molassodon is differentiated in only 48/~4800 nearly fixed core genes and over 70 accessory genes, and 3) that a significant portion of the genotype may have come from outside of *V. parahaemolyticus.* Having characterized the genotype, we moved on to understanding the Molassodon phenotype and the connection of coadapted genes with it.

Molassodon strains do not swarm despite inducing lateral flagella

Our strongest clue to a tractable phenotypic difference between Molassodon and Typical ecospecies was motility, based on the strong differentiation of lateral flagella genes (**Fig. 1A**). In *V. parahaemolyticus*, lateral flagella are environmentally regulated and enable swarming motility on surfaces, in contrast to the constitutively expressed polar flagellum that mediates swimming in liquids. Closer examination revealed that both structural components and multiple regulators of the lateral flagellar system are differentiated in Molassodon strains (**Fig. 2A**). However, the most differentiated core gene based on the number of SNP differences encodes LafK, a key regulator of lateral flagella. This suggested there might be not only a functional divergence in Molassodon lateral flagella, but also a difference in how they are regulated in response to the environment.

Laboratory motility assays revealed a striking phenotypic divergence between Typical and Molassodon strains. We compared swimming (through soft agar, driven by polar flagella) and swarming (migration on a solid agar surface driven by lateral flagella) in natural Molassodon (n=6) and Typical (n=9) isolates. Molassodon strains formed modestly larger motility halos than Typical strains in 0.3% swim agar (**Fig. 2B**, *left*), but strikingly, showed no surface migration on swarm plates, and instead grew as apparently immobile macrocolonies (**Fig. 2B**, *right*). Despite the absence of swarming in Molassodon, RNA-seq confirmed surface induction of lateral flagella genes, albeit at ~10-fold lower levels than in Typical strains (**Fig. S8A**; **Table S8**). Molassodon strains could also still assemble lateral flagella when grown on surfaces despite lacking macroscopic movement (**Fig. 2C**).

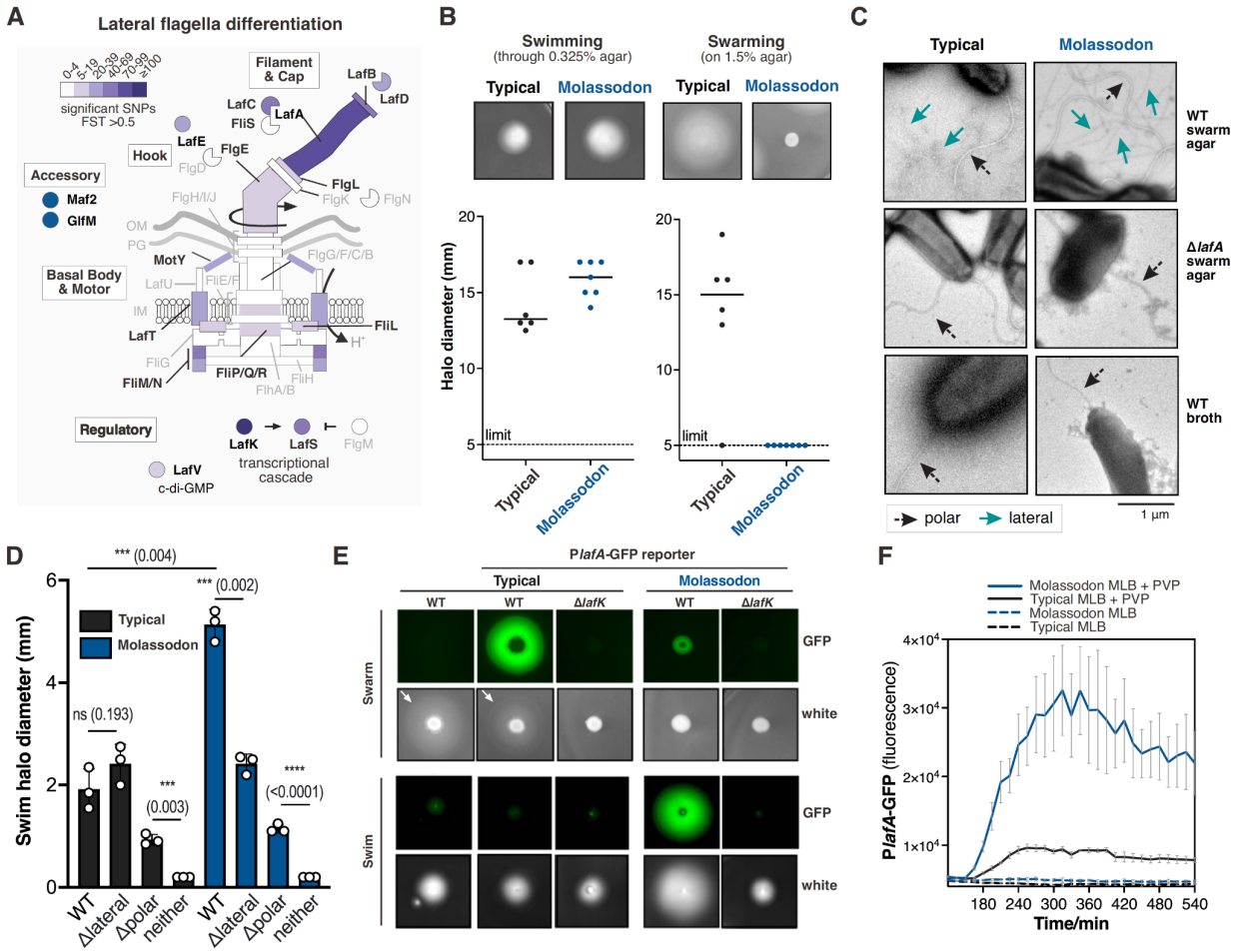

Figure 2. Molassodon uses lateral flagella for swimming in viscous media rather than swarming. (A) Schematic of lateral flagella differentiation in Molassodon strains. Differentiated: F_{st} ≥0.5 and -log10(p) ≥10; Molassodon vs Typical GWAS. **(B)** *Top*: Representative swim/swarm halos for Typical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

Molassodon uses lateral flagella for swimming through viscous conditions

We hypothesized that Molassodon strains might use their differentiated lateral flagella for an alternative function. To first confirm that Molassodon lateral flagella could in fact drive movement of any sort, we constructed mutants expressing only polar or only lateral flagella and measured migration in more permissive 0.25% soft agar. Both Molassodon and Typical strains expressing only lateral flagella (Δ*fliS*; VP2254, polar flagellin chaperone) could still form small swim halos, which were abolished by additional deletion of the major lateral flagellin gene *lafA* (**Fig. 2D**). These results demonstrate that Molassodon lateral flagella should be functional to drive motility.

This experiment unexpectedly revealed a "positive" Molassodon motility phenotype (*i.e.*, stronger than in Typical and not a "defect") that was driven by lateral flagella. First, we noticed that Molassodon strains formed larger motility halos than Typical strains in soft agar, an effect that was amplified and also significant in the more permissive, lower agar concentration (0.25%; **Fig. 2D**) compared to what we observed in our first assays (**Fig. 2B**). Strikingly, deletion of *lafA* reduced Molassodon swimming by half compared to the wild-type Typical strain, but did not affect Typical strain migration. This is consistent with the expected minimal lateral flagella expression during swimming in Typical *V. parahaemolyticus*. Moreover, it provides evidence that Molassodon might uniquely induce and use its lateral flagella in these conditions, conferring a motility advantage.

To confirm that Molassodon induces lateral flagella during swimming, we generated a transcriptional GFP reporter for the promoter of *lafA* for each ecospecies (**Fig. S8B**). As expected, Typical strains induced the *lafA* promoter only on swarm agar, whereas Molassodon activated it while swimming (**Fig. 3E**). RNA-seq of three strains per ecospecies taken from swim agar confirmed higher expression of all lateral flagella genes in Molassodon compared with Typical in this condition (**Fig. S7C**; **Table S9**), which is in line with the Δ*lafA* swimming defect in only Molassodon (**Fig. 2D**).

What could be the specific signal resulting in differential induction of lateral flagella between the ecospecies? Surface sensing and viscosity can trigger lateral flagella expression via inhibition of polar flagella rotation in *V. parahaemolyticus* [35], so we hypothesized that this could be one environmental trigger. To test this, we next measured the effect of polyvinylpyrrolidone (PVP), a polymer that increases viscosity, on expression of the *lafA* reporters. We found that Molassodon induced *PlafA* more strongly than Typical strains in response to increased viscosity (**Fig. 2F**; **Fig. S7D**). Together, these results reveal the first laboratory phenotype for a bacterial ecospecies associated with differentiated genes, and suggest that Molassodon might uniquely induce and use lateral flagella for swimming in viscous environments.

Functional linkage of most differentiated genes to lateral flagella-dependent swimming

Having identified, at least in the laboratory, a "positive" phenotypic difference for Molassodon that is linked to lateral flagella (swimming in viscous media), we hypothesized that we could leverage it to assay if other Molassodon-differentiated genes might also be linked to this behavior. Potentially, the other differentiated genes could be 1) hidden essential motility factors, or 2) genes that are co-

expressed under the same condition because they are part of a broader, emergent phenotype representing a complex adaptation involving most of the multilocus genotype. To test these two possibilities, we used Tn-seq to identify genes required for Molassodon and Typical swimming ($\log_2$FC < 1 vs. shaken media) and our RNA-seq dataset from the swim condition to identify genes co-expressed with lateral flagella ($\log_2$FC > 1 vs. LB) in each ecospecies (**Fig. 3A**). As a control, we also measured phenotypes and expression in the Typical strain during swarming (**Fig. S9A**).

Of the 116 Molassodon genes that we could assay with both Tn-seq and RNA-seq, 76 (66%) were either required for swimming or induced in the swim condition where only Molassodon uses lateral flagella (39 induced, 56 required, 19 overlapping; **Fig. 3A**; **Table S10**; **Fig. S9**). In contrast, few differentiated core genes were induced or required by Typical in the swim condition (**Fig. 3A**; **Table S11**). As a control, we confirmed that Typical lateral flagella genes were required for swarming (**Fig. S9A**; **Table S12**). Overall, Molassodon genes were highly enriched in both swim-induced and swim-required sets (padj = 5.36×10^{-16}, NES = 3.36; padj = 3.82×10^{-12}, NES = −2.68), indicating that this swimming phenotype, although tested in the lab, might encompass most Molassodon-differentiated genes. It follows that we may have potentially recapitulated part of the ecological scenario where Molassodon genes together provide an adaptive advantage.

To understand why the other required/co-expressed Molassodon genes might be linked to motility, we examined their functions. First, we noted that several differentiated genes in the LPS ("O") locus, including genes for legionaminic acid (Leg) sugar biosynthesis, showed strong Tn-seq phenotypes (**Fig. S10A**). A block of K-locus genes, previously suggested to be gained via insertion event from other species [34], also had a strong Tn-seq phenotype. Synteny and structural similarity analysis suggested that these genes are homologous to a recently described marine bacterial S-layer and its dedicated Type II secretion system [36] (**Fig. S10B-D**). 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 [37]. We confirmed via SDS-PAGE analysis that Molassodon strains express a highly abundant protein that disappears upon deletion of the putative S-layer protein gene, *slr4* (**Fig. S10E**), although we have not yet visualized an S-layer. Finally, this region also encodes a distinct ABC transporter-dependent capsule, unlike the Wzx/Wzy-dependent version of Typical strains [34], which also mildly affected swimming (**Fig. S10A**).

Tn-seq phenotypes for many of the O:K locus genes such as those encoding Leg biosynthesis (**Fig. 3A**) at first suggested altered cell envelope/surface characteristics might indirectly affect swimming. However, we did not see the same impact on Typical swimming (or swarming) (**Fig. 3A; Fig. S9A**). This led us to investigate a direct link between other differentiated genes and lateral flagella.

Two uncharacterized Molassodon accessory genes encoded downstream of *lafA* (now named *maf2* & *glfM*) showed a strong swimming defect in the Tn-seq experiment (**Fig. 3A; Fig. S10A**). Based on homology, predicted complex formation, and co-inheritance with the Molassodon *lafA* allele, these two genes might encode a Maf2-GlfM flagellin glycosylation system as was recently described in *Shewanella* [38] (**Fig. 3B & 3C**). Maf2 and GlfM likely act together to glycosylate Molassodon LafA with a sialic-acid–derived sugar. Consistent with this, the Molassodon OL2/OL5/OL11 LPS islands encode a putative pathway for the biosynthesis of the sialic acid legionaminic acid (Leg) (**Fig. 3D; Fig. S7B**),

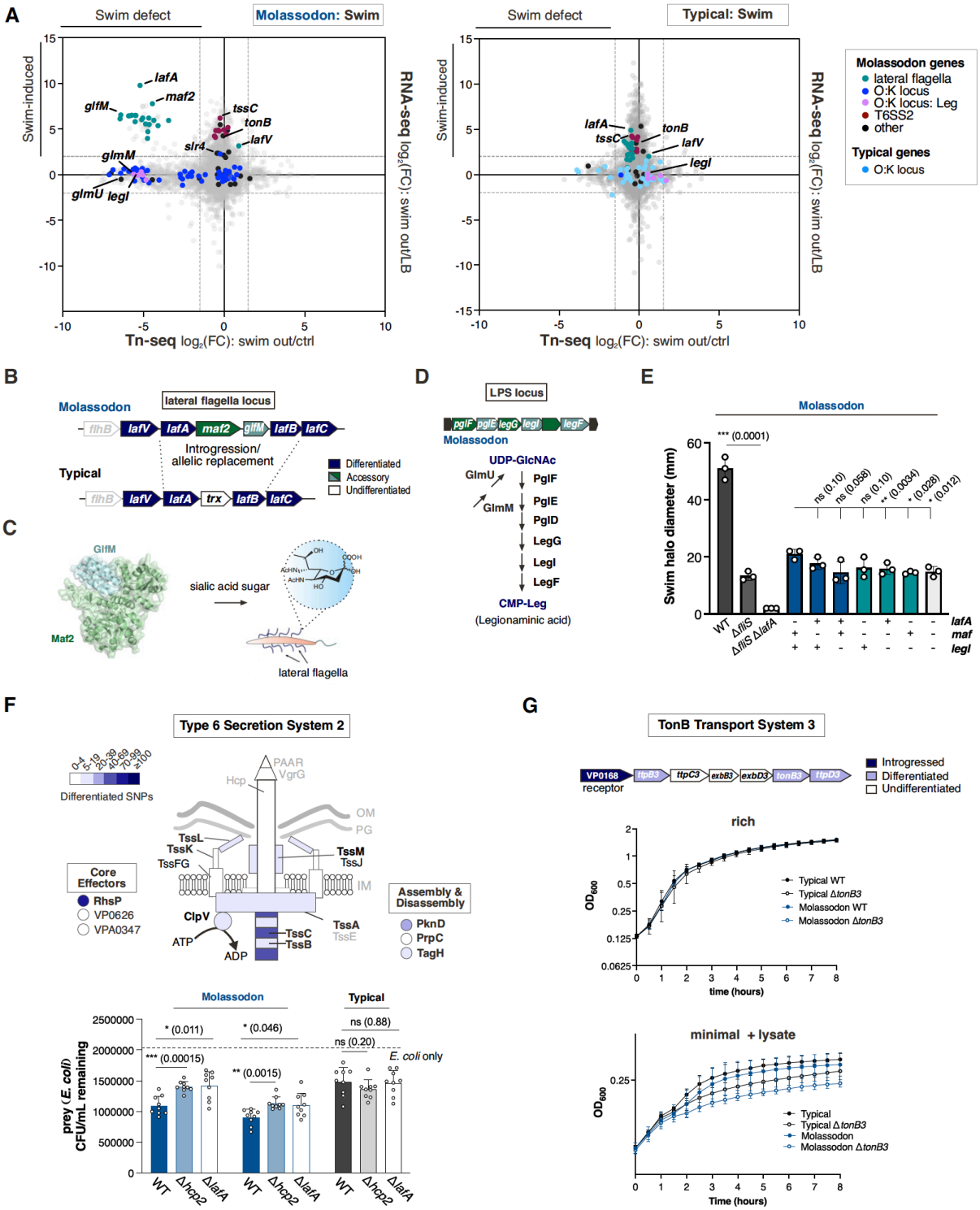

Figure 3. Validation and refinement of the Molassodon laboratory phenotype with Tn- and RNA-seq. (A) Tn-seq and RNA-seq profiles for Molassodon and Typical strains during swimming. Tn-seq: RIMD 2210633 (Typical) and DFS-0010 (Molassodon), $\log_2$FC Swim halo vs. Swim control. RNA-seq: mean of three 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. S9B** and **Tables S10-11**. **(B)** Genomic

context of the Molassodon accessory lateral flagella genes *maf2-glfM* and related region in non-introgressed strains. **(C)** Predicted complex formed by Maf2-GlfM and role of the proteins in flagellin glycosylation. The complex was predicted with Alphafold Multimer at Colabfold [39,40] and visualized with Pymol. **(D)** Molassodon genes encoding a putative Leg (legionaminic acid) biosynthesis pathway. The pathway is based on [41]). CMP: cytidine monophosphate. **(E)** 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. *fliS*: polar flagella chaperone. **(F)** *Top:* Schematic of T6SS2 with differentiated components indicated. Differentiated: $F_{ST} \geq 0.5$ and -log10(p) ≥10); Molassodon vs Typical GWAS. *Bottom:* 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. **(G)** *Top:* Schematic of differentiated genes encoding a TonB/ExbBD transporter and receptor in Molassodon. *Bottom:* Growth curve analysis of Molassodon and Typical strains (n=3 each) and isogenic tonB3 (VP0163) deletion strains in rich (MLB) or minimal medium (M9) with prey (*E. coli*) lysate is provided as a nutrient source.

and O2 serotype LPS carries a Leg derivative [42]. Molassodon *lafA*, *maf2*, and *legI* appear to have been imported from outside of *V. parahaemolyticus*, being closer to that found in *V. antiquarius* (**Fig. S11A**).

All six Leg pathway genes, along with two Molassodon-differentiated core genes feeding into this pathway (*glmM*, *glmU*), had Molassodon-specific Tn-seq phenotypes in the swim condition, but did not affect Typical swimming or swarming (**Fig. 3A, Figs. S9A & S10A**). Single and combinatorial mutants of *lafA*, *maf-glfM*, and *legI* phenocopied each other, producing swim halos similar to Δ*lafA*, which provides support that these genes act in a shared pathway affecting Molassodon lateral-flagella-dependent swimming (**Fig. 3E**). Therefore, a functional connection to lateral flagella explains unexpected Tn-seq phenotypes for several additional genes, including accessory genes specific to Molassodon and core genes encoded outside the lateral flagella loci. This also suggests that Leg biosynthesis and LPS modification preceded the acquisition of flagellin glycosylation by Maf during Molassodon evolution, providing a first into how the ecospecies may have assembled.

Type VI-mediated killing and nutrient uptake are also functionally linked to Molassodon lateral flagella

Many Molassodon genes did not have a swimming phenotype. However, many were co-expressed with lateral flagella during swimming, suggesting that the natural adaptation might encompass more than just motility, and that understanding these co-expressed genes might help us refine the natural phenotype. Two co-expressed clusters stood out: those encoding the T6SS2 Type VI secretion system (*e.g.*, *tssC*) and the TonB3 active transport system (**Fig. 3A**; **Fig. S10A**).

T6SS are contact-dependent killing systems, often directed towards other bacteria. Genes encoding several T6SS2 components are differentiated in Molassodon, including structural genes, the effector RhsP, and assembly/disassembly factors (**Fig. 3F**, *top*). We reasoned that the connection between lateral flagella and antagonism might stem from the contact-dependent nature of T6SS. Motility has been shown to enhance T6SS activity in other bacteria [43–45], and we hypothesized that this might underlie coadaptation of these two functions in Molassodon. To provide evidence that lateral flagella-

driven motility might impact the efficacy of Molassodon T6SS2, we took advantage of PVP-specific induction of *lafA* specifically in Molassodon (**Fig. 2F**) and tested T6SS2-dependent killing of prey *E. coli* under these conditions. Although killing in PVP was modest (~0.5 log) compared to agar-based assays [46], deletion of *lafA* significantly increased prey survival only in Molassodon strains (**Fig. 3F**, *bottom*). While these are simplified laboratory observations, taken together with co-evolution of T6SS2 and lateral flagella and observations in other bacteria [43–45], these results provide experimental evidence that Molassodon T6SS2 activity could be dependent on lateral flagella in viscous conditions. This suggested to us that in a natural setting, the broader adaptation of Molassodon could be related to active antagonism.

Recent studies have also shown that T6SS-mediated killing by *Vibrio* can provide nutrients from lysed cells, and also detected a negative correlation between T6SS genes and metabolic diversity across species [8]. We therefore investigated whether antagonism, driven by motility, could be a "hunting" strategy. Interestingly, the second co-expressed, differentiated gene cluster encodes the TonB3 nutrient uptake system (**Fig. 3G**, *top*), one of three TonB systems in *V. parahaemolyticus* (TonB1–3; *V. vulnificus* nomenclature). TonB systems transduce energy from the proton-motive force to drive active transport across the outer membrane, and recent studies have shown a functional link between T6SS, such uptake systems, and ion acquisition [47–50]. This led us to hypothesize that coadaptation of lateral flagella, T6SS2, and TonB3 might reflect a functional linkage in the uptake of specific resources following T6SS-mediated lysis.

To test whether TonB3 could potentially facilitate nutrient acquisition from lysed bacterial prey, we grew Molassodon and Typical (n=3) wildtype and Δ*tonB3* (VP0163) strains in either rich medium (MLB) or M9 minimal salts with *E. coli* lysate as the sole nutrient source. While WT strains showed no clear growth differences under these conditions (**Fig. S12**), all Δ*tonB3* mutants reached lower densities than their respective wildtypes (**Fig. 3E**). These results suggest that TonB3 could contribute to uptake of lysate-derived nutrients in both ecospecies, although the specific substrates and ecological context remains to be demonstrated.

Nonetheless, together these experiments provide initial evidence for a functional link between three coadapted pathways in Molassodon—lateral flagella–driven motility, Type VI secretion, and nutrient uptake—supporting functional coherence of the genotype and providing insight for a potential ecological role.

Convergent association of lateral flagella and TonB in a putative intermediate genotype

We noticed that some Typical *V. parahaemolyticus* strains (15-25% in our dataset) also encode a *maf2* homolog (**Fig. 1C**; **Table S5**). However, this version of *maf* is divergent from the Molassodon version, and seems to be more closely related to a *V. antiquarius* version and is possibly a more recent import (**Fig. 4A**; **Fig. S11**). As Maf2 is functionally linked to lateral flagellin, we next compared *lafA* sequences between Molasson and Typical *maf-*/*maf*+ strains. This suggested there are three types of *lafA,* which follow *maf* genotype: two associated with Molassodon or Typical *maf2* and potentially imported from outside of *V. parahaemolyticus*, and one restricted to *maf- strains* (**Fig. 4B**).

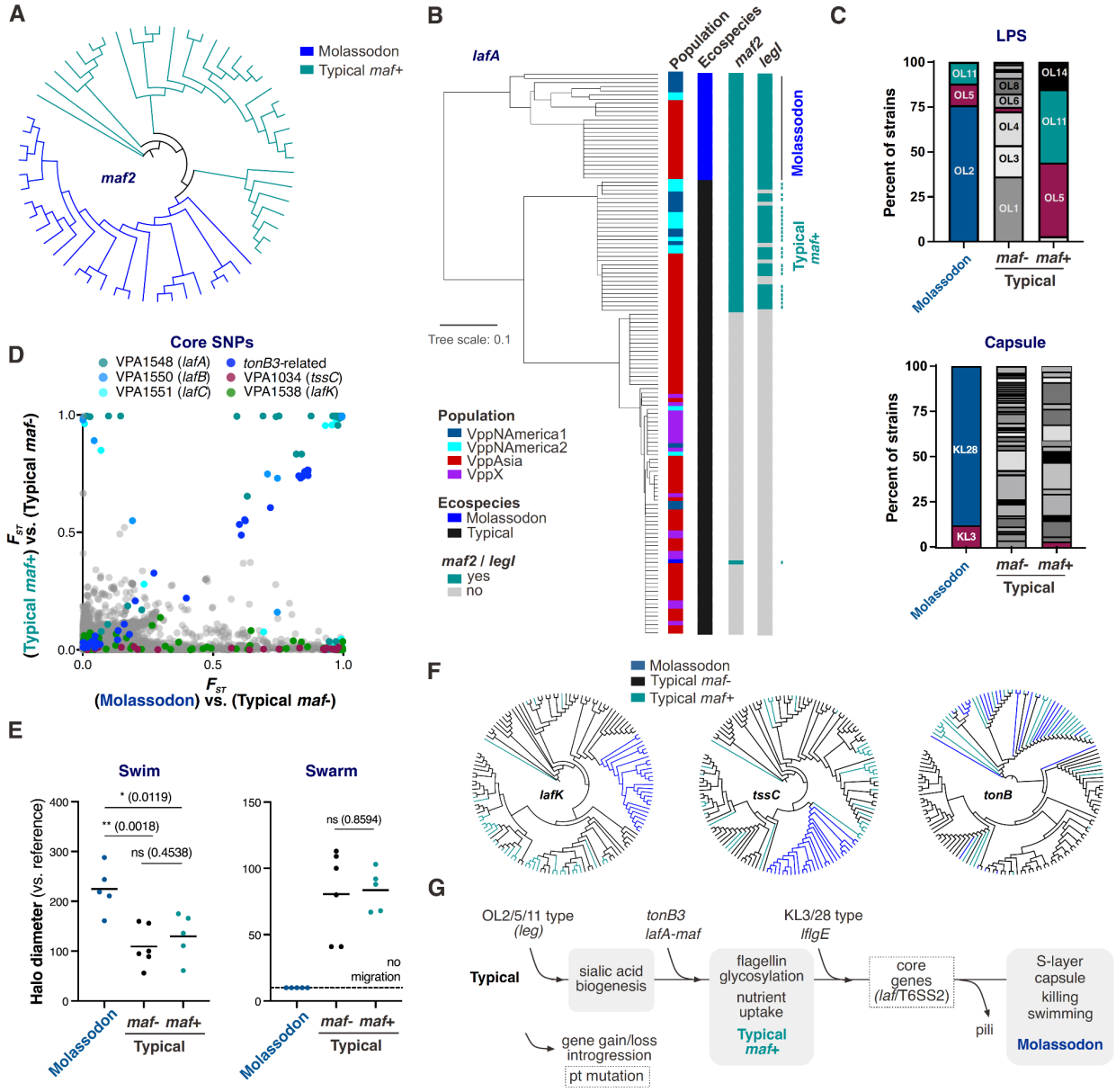

Figure 4. Convergent coadaptation after *maf* introgression. (A) Neighbour-joining tree based on *maf2* homologs identified by BLAST. For branch lengths, see **Fig. S11B. (B)** Neighbour-joining phylogenetic tree based on *lafA* for a subset of strains with presence of "*maf*" (*maf2-glfM*) indicated. **(C)** 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**. **(D)** Core genome SNP F_{ST} (fixation index) correlation for (Typical *maf+*) vs. (Typical *maf-*) plotted against corresponding values for (Molassodon) vs. (Typical *maf-*). Key functions identified in this and previous [10,30] work are highlighted. **(E)** Swimming and swarming of wild-type strains from different ecospecies/*maf* genotypes (Molassodon, Typical *maf-*, Typical *maf+*). Symbols represent the mean of two independent replicates for each strain. Diameters are normalized to the reference strain (RIMD 2210633) tested on the same motility plate. ns: not significant, Student's *t*-test. **(F)** Neighbour-joining trees based on *lafK* (lateral flagella), *tssC* (T6SS2) and *tonB3*. or branch lengths, see **Fig. S11B. (G)** Evolutionary model for assembly of Molassodon.

Above, we noted limited predicted serotype diversity in Molassodon strains (**Fig. 1E**), and we observed a similarly limited diversity for Typical *maf*+ strain O types (**Fig. 4C**), with two types overlapping with Molassodon (OL5 and OL11) and a single type (OL14) shared with only one Typical *maf*- strain. K-locus (capsule) types were as diverse as Typical *maf*- strains, demonstrating that Typical *maf*+ do not overlap completely with Molassodon. However, consistent with an intermediate genotype, we also noticed that most *maf*+ Typical strains are in an intermediate position between Molassodon and Typical *maf*-strains in a tree based on Molassodon-differentiated core genes (**Fig. 1C**). We hypothesized that this additional imported *lafA-maf* genotype, which also shows association with only a few O-locus types, might represent a convergent, functionally overlapping, assembly of a coadapted genotype also involving lateral flagella.

We next searched for additional potentially epistatic relationships shared with Molassodon via GWAS. This revealed only a small number of core genome regions associated with *maf2* in Typical strains (**Fig. S13A,B)**. However, the two most striking regions overlapped with Molassodon: a proportion of SNPs in *lafA* and the downstream genes *lafB* and *lafC*, as well as the region encompassing the *tonB3* genes. We also noted that the degree of fixation (F_{ST}) for many SNPs in *lafABC* and *tonB3* correlated between the intermediate strains and Molassodon, although the *tonB* variants have not reached full fixation in either population (F_{ST} >0.9) (**Fig. 4D**). Many SNPs were uniquely differentiated in Molassodon, such as those in other core lateral flagella genes (e.g., *lafK*) and in those encoding T6SS2 (e.g., *tssC*) (**Fig. 4C,D**). In the accessory genome, the distribution of many gene variants was also correlated for Molassodon and the intermediate, such as *leg* genes (**Fig. S12C**), while again others were restricted to Molassodon, such as those in the K-locus (including the putative S-layer), consistent with serotype predictions (**Fig. 4C**).

These results suggested that variation in lateral flagella and nutrient uptake have been independently associated within subpopulations during *V. parahaemolyticus* evolution. However, while present-day *maf*+ intermediate strains seem to recapitulate 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. 4D; Fig. S13A,B**). Compared to Typical *maf*+ strains, Molassodon shows much more extensive divergence at core lateral flagella loci, with 233 SNPs across 15 genes, versus only three differentiated genes between *maf*– and *maf*+ Typical strains (*lafA*, *lafB*, and *lafC*). Notably, the LafK regulator has not diverged from *maf*- strains in the intermediate strains (**Fig. 4E**; **Fig. S11B**), suggesting that distinct environmental regulation of lateral flagella has not yet evolved. Consistent with this, all Typical *maf*+ strains we tested (n=5) swarm on agar and show no obvious differences in swimming behaviour from *maf*- strains (**Fig. 4F**). However, introducing *maf* from Molassodon into a *maf*– *leg*+ (but not *leg*–) Typical strain altered swarming (**Fig. S14A**), indicating that *maf* acquisition could affect motility and potentially enable access to a new niche. Finally, the key differentiation at T6SS2 is also restricted to Molassodon, suggesting that Typical *maf*+ strains have not coadapted antagonism with lateral flagella-driven motility and nutrient uptake.

Based on our genetic characterization of the Typical *maf*+ strains and motility assays, we propose a working model for how the Molassodon genotype may have arisen (**Fig. 4G**). It appears that acquisition of lateral flagellin glycosylation and potentially other structure-function changes in LafA from other *Vibrio* species in strains that already encode genes that can provide a sugar substrate for

Maf2-GlfM can affect motility behaviour, potentially allowing access to a new niche. According to this model, as ecological differentiation continued, additional strategies (nutrient acquisition, or even "hunting") co-evolved, as well as regulatory differences that altered behaviour. Finally, extensive cell surface remodeling occurred, due to differences in physical host/environmental interactions. This is likely to have included, at some point, protection from predation from other bacteria or phages (e.g., via the S-layer, different capsule, cellulose layer, changes in flagellin glycosylation, loss of a putative pilus (VPA0713–31; **Fig. 1C**)).

Discussion

Signals of natural selection provide insight into ecological differentiation in bacteria. Here, we showed that *Vibrio parahaemolyticus* Molassodon forms an ecospecies defined by ~100 coadapted genes that remain fixed by natural selection, despite rampant genome-wide recombination. The structure is similar to the *Helicobacter pylori* Hardy ecospecies but with substantial contributions from accessory genes and introgression. We provide the first experimental evidence of ecospecies functional coherence by linking most differentiated genes to enhanced swimming in viscous media. Based on their ecological selection and functional linkage, we hypothesize that this multilocus genotype supports a hunt–kill–devour strategy. Finally, we identify a partially convergent genotype, which provides additional support for ecological selection and suggests there are recurring adaptive axes in *V. parahaemolyticus*.

Molassodon is an ecospecies defined by a separate gene pool

First, we showed that we could robustly define the Molassodon genotype as a well-demarcated set of 48 core and 73 (present) accessory genes (**Figs. 1A & 1C**). The clean, sharp peaks we obtain in our differentiation scan (**Fig. 1A**) allows us to identify which loci are under selection 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**). The genotype is found on multiple otherwise unrelated *V. parahaemolyticus* backgrounds, including three out of four genetically differentiated oceanic populations, implying that it encodes a stable evolutionary strategy, applicable in many locations.

Our study of Molassodon highlights a role for introgression in incorporation of new adaptive traits during ecological differentiation. For example, we identified two independent imports of neighboring *lafA* and *maf2*, one leading to Molassodon and the other to a much simpler coadapted genotype, which we hypothesize represents an earlier stage of functional divergence. We functionally linked these *maf* imports to additional imported genes required to generate the legionaminic acid sugar substrate for glycosylation, as well as specific alleles of LafA, the target for glycosylation. As legionaminic acid could also be used to modify LPS independent of *maf* [42], this provides a feasible pathway for stepwise assembly of part of the complex adaptation (**Fig. 4G**).

The distinct [34] capsule locus of Molassodon is also a reservoir for many other introgressed traits that affect swimming directly or indirectly, such as a putative S-layer. So far, *Vibrio* expressing an S-layer has not yet been described, although *Vibrio* spp. encoding a newly identified S-layer in marine species have been reported [36]. This putative S-layer, as well as the distinct capsule of Molassodon, may provide protection during bacterial antagonism/predation or from phage adsorption. Alternatively, it may confer specific stress tolerance or adhesion properties relevant to the niche of Molassodon [37].

Our study of Molassodon thus provides additional evidence for a pan-*Vibrio* gene pool of functionally compatible motility-related and cell envelope genes, in addition to metabolic genes as previously reported [51]. We propose that this toolbox facilitates continuous evolutionary innovation in ecological strategies in the genus.

We have previously identified ecospecies in three other bacterial species, albeit without functional validation. The phenotypes appear to be completely unrelated between the species, but involve facets that are central to the biology and the niche of each. Putative ecospecies in *Streptococcus mitis* and *S. oralis* appear to have convergent evolution in cell envelope/division-related genes [29], while *H. pylori* ecospecies have differentiated in genes related to cell adhesion, interaction with the immune system, and acid tolerance. Amongst eukaryotes, cichlid fish are known to speciate rapidly thanks to flexible jaw morphology [52] 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 [29] and may precipitate speciation [10].

Comparison of bacterial ecospecies to other genetic structures

Ecotypes have been defined theoretically as ecologically distinct, genetically cohesive units maintained by periodic selection [18] and operationally as phylogenetic groups with a distinct sampling distribution [20]. Bacterial ecospecies are not genetically cohesive at a genomic level, since outside the ecospecies-defining loci they share the entire diversity of the species, and for this reason they do not necessarily cluster together based on phylogenetic methods. Indeed, Molassodon is much older than any identifiable phylogenetic lineage in *V. parahaemolyticus* and is so far shared between at least three of the four identified geographic populations. Furthermore, unlike ecotypes, ecospecies can be identified based on genetic differentiation alone, without any ecological context.

The early-stage differentiated groups described by Shapiro *et al.* in *V. cyclotrophicus* [21] resemble the early stages of differentiation such as seen in the intermediate Typical *maf*$^{+}$ genotype that we identified and that we have argued is a plausible precursor to ecospecies. Shapiro and colleagues emphasized the role of reduced recombination between demes in maintaining cohesion, while selection drives divergence at ecologically relevant loci. There is no sign of barriers to recombination between the *V. parahaemolyticus* ecospecies, since the great majority of the genome is undifferentiated. Therefore, in our context, substantial functional differentiation appears to have preceded meaningful barriers to genetic exchange. Nevertheless, as has been emphasized repeatedly in the eukaryotic speciation literature [53,54] that barriers to gene flow can foster the

evolution of functional differentiation and speciation and are likely to be a widely important force in prokaryotes also.

The genetic structures identified to date that most closely represent bacterial ecospecies are found in eukaryotes, such as stickleback fish, which repeatedly recolonize similar environments. Each time a new river environment is colonized by marine sticklebacks, 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]. Thus, these species have become models for ecological genomics. Our peaks (**Fig. 1A**) are narrower to those seen in sticklebacks, or other eukaryotic model systems such as butterflies, which typically contain many genes [2,55]. Our sharp genetic peaks increase accuracy in every stage of functional inference.

Functional linkage of Molassodon genes suggests an emergent hunt–kill–devour strategy in viscous environments

We were able to identify a laboratory phenotype associated with the Molassodon differentiated gene set, providing first experimental support for the functional relationship between the genes and allowing us to make hypotheses about the ecological relevance of the genotype. Molassodon exhibits enhanced swimming in viscous conditions in the lab, but loss of swarming behaviour. We could show that many differentiated genes are specifically expressed or required for enhanced swimming, along with lateral flagella. Some of this linkage was explained by demonstrating that several Molassodon genes cooperate to generate glycosylated lateral flagella. Some of the Tn-seq phenotypes could be due to a more indirect effect on swimming, such as changing surface characteristics, while others could have direct, uncharacterized roles in motility.

While expression does not necessarily confirm functional interaction, we observed that T6SS2 and TonB appear to be co-expressed during swimming. Based on previous experimental and genomics in other bacteria linking T6SS and TonB-mediated micronutrient uptake [47–50], we hypothesized that they might cooperate in killing bacterial prey and mediating uptake of lysis products. We were able to demonstrate that lateral flagella do specifically affect T6SS-mediated killing, and that TonB confers growth when lysate is provided as the sole nutrient source, under laboratory conditions. We do not know the nutrient that TonB is transporting in either ecospecies, but high sequence divergence could indicate functional specialization between the two (**Fig. S6B**). Nutrients such as iron or vitamin B12 are scarce and/or microbe-associated in marine environments, and phage-mediated lysis has been proposed to provide a source of B12, a common substrate for TonB transporters [56]. In *Klebsiella*, clusters of ABC-type ion transporters and *tonB* homologues have been found encoded adjacent to T6SS genes, suggesting that the nutrient uptake-T6SS link could be a more universal ecological strategy for free-living bacteria.

We have not linked the hunt kill and devour aspects in a single experiment, and it might be difficult to do so without better mimicking the natural setting where these functions interact. However, our inference is supported by combined evidence from coadaptation and coregulation, and our laboratory experiments suggest how they could cooperate in a coadapted, interacting functional module rather than as individual traits. The natural microenvironment where the genes together

provide an adaptive advantage is unknown, but could resemble, *e.g.,* mucus layers of host species (fish, invertebrates) where they must compete for colonization space and nutrients, diffusive boundaries surrounding particles, or phytoplankton exopolysaccharides. Currently, around a third of ecospecies genes do not have a laboratory 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 [57], 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 a mechanism to actively trap T6SS prey with a "grappling hook" has been reported [58]. Our experimental and coadaptation results provide evidence for flagella-driven predation, whose effectiveness could be enhanced in a viscous environment by limited nutrient diffusion [59]. We have also discovered an intermediate genotype which also links nutrition and motility, highlighting their intimate link in this species.

Ecospecies as a tool for studying complex adaptations

Because bacterial ecospecies can be identified based on genetics alone, without the need for ecological clues, our approach is a particularly pure 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 [17,60]. Here, we demonstrate how molecular genetics and global approaches such as RNA-seq and Tn-seq can be leveraged to gain insight into ecospecies and other complex adaptations.

We hypothesize that the natural phenotypic difference of Molassodon is related to hunting, killing, and devouring in viscous environments but have left field investigation to future studies. Our knowledge of where and how *Vibrio* swim, swarm and hunt in the natural environment is currently extremely limited. Because Molassodon and Typical strains and intermediate *maf*+ genotypes coexist (**Fig. S14B**) 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) [63], followed by genome assembly with Flye (v2.9-b1768) [64].

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 [65]. Genomes were annotated using Prokka version 1.14.6 [66] and a pangenome was built using Panaroo version 1.3.3 [67].

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 [68] 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 [68] 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 [69], 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} [70] was computed using vcftools [71] 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 [67] 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 [72]. 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 [73]. Trees were visualized using iTOL [72].

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 [30], 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 [27]. 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 [31,32]. Synteny analysis was performed with webflags (www.webflags.se) [74]. Structures of Slr4 homologues and protein complexes (Maf2 + GlfM) were predicted with AlphaFold2 or Alphafold multimer at Colabfold [39,40]. Structures were visualized with Pymol version 3.0.2. RNA-seq coverage was displayed with Integrated Genome Browser [75].

Bacterial lab strains and culture conditions. Bacterial strains are listed in (**Table S7**). Most *V. parahaemolyticus* wild-type 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. 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: 1% (w/v) tryptone, 0.5% yeast extract, 1% NaCl, Sangon) plates or MLB (LB with 3% NaCl final concentration) 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 [76]. For swarming, we used either previously described BHI agar (1.5%) [76], 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 [77], as described previously [14,76]. 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) [78] 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 [79] 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 [80] 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 [81]. 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 [82]. Reads were then mapped to the corresponding reference genome using Segemehl version 0.3.4 with default parameters [83]. FeatureCounts version 2.0.3 was used for gene expression quantification with parameters "-B -C -f -p -s 1 -O --fraction --countReadPairs --ignoreDup –primary" [84]. Differential expression was performed using R package DESeq2 version 1.42.1 using the output from featureCounts without normalization [85]. 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 [86]. Gene set classifications include GO terms [87], KEGG pathways, KEGG KO [88], COG categories [89], 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 [90] and Julia. Plots were generated using Julia package Makie.jl [91].

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[92] 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 [76] 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 [93], 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) [94]. Subsequently, the read counts of each insertion site were tallied using the TraDIS toolkit [95]. 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 (Sangon) with 2% (w/v) NaCl (final concentration) 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.

Total protein analysis. Protein expression in *V. parahaemolyticus* was performed by SDS-PAGE of whole-cell lysates. Strains were grown in MLB at 30°C until log phase, and then cells were harvested by centrifugation at 13,000 rpm. Cell pellets were resuspended in 100 µl of 1× protein loading buffer (62.5 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol, 10% (v/v) glycerol, 2% (w/v) SDS, 0.01% (w/v) bromophenol blue) and boiled for 8 min. Next, 0.1 OD_{600} of cells were loaded per lane on 12% SDS-polyacrylamide gels. After electrophoresis, protein bands were visualized by staining with Coomassie Blue (Sangon).

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. *et al.* Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat. Commun.* **7**, 11693 (2016).
2. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
3. Tourrette, E. *et al.* An ancient ecospecies of *Helicobacter pylori*. *Nature* (2024) doi:10.1038/s41586-024-07991-z.
4. Thorell, K. *et al.* The *Helicobacter* pylori Genome Project: insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. *Nat. Commun.* **14**, 8184 (2023).
5. Broberg, C. A., Calder, T. J. & Orth, K. *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes Infect.* **13**, 992–1001 (2011).
6. Grognot, M., Nam, J. W., Elson, L. E. & Taute, K. M. Physiological adaptation in flagellar architecture improves *Vibrio alginolyticus* chemotaxis in complex environments. *Proc Natl Acad Sci USA* **120**, e2301873120 (2023).
7. Speare, L. *et al.* Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proc Natl Acad Sci USA* **115**, E8528–E8537 (2018).
8. Stubbusch, A. K. M. *et al.* Antagonism as a foraging strategy in microbial communities. *Science* **388**, 1214–1217 (2025).
9. Freitas, C., Glatter, T. & Ringgaard, S. The release of a distinct cell type from swarm colonies facilitates dissemination of *Vibrio parahaemolyticus* in the environment. *ISME J.* **14**, 230–244 (2020).
10. Cui, Y. *et al.* The landscape of coadaptation in *Vibrio parahaemolyticus*. *eLife* **9**, (2020).
11. Homma, M., Nishikino, T. & Kojima, S. Achievements in bacterial flagellar research with focus on *Vibrio* species. *Microbiol. Immunol.* **66**, 75–95 (2022).
12. Kocher, A., Krause, J. & Spyrou, M. A. Insights into infectious diseases through ancient pathogen genomics. *Nat. Rev. Microbiol.* (2025) doi:10.1038/s41579-025-01259-7.
13. Espejo, R. T., García, K. & Plaza, N. Insight Into the Origin and Evolution of the *Vibrio parahaemolyticus* Pandemic Strain. *Front. Microbiol.* **8**, 1397 (2017).
14. Yang, C. *et al.* Wave succession in the pandemic clone of *Vibrio parahaemolyticus* driven by gene loss. *Nat. Ecol. Evol.* (2025) doi:10.1038/s41559-025-02827-z.
15. Maier, L. *et al.* Integrating research on bacterial pathogens and commensals to fight infections-an ecological perspective. *Lancet Microbe* **5**, 100843 (2024).
16. Diard, M. & Hardt, W.-D. Evolution of bacterial virulence. *FEMS Microbiol. Rev.* **41**, 679–697 (2017).
17. Shapiro, B. J. & Polz, M. F. Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol.* **22**, 235–247 (2014).
18. Cohan, F. M. & Perry, E. B. A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* **17**, R373-86 (2007).
19. Cohan, F. M. & Koeppel, A. F. The origins of ecological diversity in prokaryotes. *Curr. Biol.* **18**, R1024-34 (2008).
20. Hunt, D. E. *et al.* Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**, 1081–1085 (2008).
21. Shapiro, B. J. *et al.* Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48–51 (2012).
22. Cordero, O. X. *et al.* Ecological populations of bacteria act as socially cohesive units of antibiotic

production and resistance. *Science* **337**, 1228–1231 (2012).
23. Trinanes, J. & Martinez-Urtaza, J. Future scenarios of risk of *Vibrio* infections in a warming planet: a global mapping study. *Lancet Planet. Health* **5**, e426–e435 (2021).
24. McCarter, L. L. Dual flagellar systems enable motility under different circumstances. *J. Mol. Microbiol. Biotechnol.* **7**, 18–29 (2004).
25. Nair, G. B. *et al.* Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin. Microbiol. Rev.* **20**, 39–48 (2007).
26. Yang, C. *et al.* Outbreak dynamics of foodborne pathogen *Vibrio parahaemolyticus* over a seventeen year period implies hidden reservoirs. *Nat. Microbiol.* **7**, 1221–1229 (2022).
27. Yang, C. *et al.* Recent mixing of *Vibrio parahaemolyticus* populations. *ISME J.* **13**, 2578–2588 (2019).
28. Yang, C., Cui, Y., Didelot, X., Yang, R. & Falush, D. Why panmictic bacteria are rare. *BioRxiv* (2018) doi:10.1101/385336.
29. Yang, C. *et al.* Macrogenetic atlas of prokaryotes. *BioRxiv* (2025) doi:10.1101/2025.02.26.640246.
30. Cui, Y. *et al.* Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen *Vibrio parahaemolyticus*. *Mol. Biol. Evol.* **32**, 1396–1410 (2015).
31. Wick, R. R., Heinz, E., Holt, K. E. & Wyres, K. L. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. *J. Clin. Microbiol.* **56**, (2018).
32. van der Graaf-van Bloois, L., Chen, H., Wagenaar, J. A. & Zomer, A. L. Development of Kaptive databases for *Vibrio parahaemolyticus* O- and K-antigen genotyping. *Microb. Genom.* **9**, (2023).
33. Chen, Y., Dai, J., Morris, J. G. & Johnson, J. A. Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol.* **10**, 274 (2010).
34. Bian, S. *et al.* Genetic Structure, Function, and Evolution of Capsule Biosynthesis Loci in *Vibrio parahaemolyticus. Front. Microbiol.* **11**, 546150 (2020).
35. Gode-Potratz, C. J., Kustusch, R. J., Breheny, P. J., Weiss, D. S. & McCarter, L. L. Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. *Mol. Microbiol.* **79**, 240–263 (2011).
36. Ali, S. *et al.* Slr4, a newly identified S-layer protein from marine Gammaproteobacteria, is a major biofilm matrix component. *Mol. Microbiol.* **114**, 979–990 (2020).
37. Fagan, R. P. & Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **12**, 211–222 (2014).
38. Unay, J., Kint, N. & Viollier, P. H. Evolution of paralogous multicomponent systems for site-specific O-sialylation of flagellin in Gram-negative and Gram-positive bacteria. *Curr. Biol.* **34**, 2932-2947.e7 (2024).
39. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
40. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *BioRxiv* (2021) doi:10.1101/2021.10.04.463034.
41. Morrison, M. J. & Imperiali, B. The renaissance of bacillosamine and its derivatives: pathway characterization and implications in pathogenicity. *Biochemistry* **53**, 624–638 (2014).
42. Hashii, N., Isshiki, Y., Iguchi, T. & Kondo, S. Structural analysis of the carbohydrate backbone of *Vibrio parahaemolyticus* O2 lipopolysaccharides. *Carbohydr. Res.* **338**, 1063–1071 (2003).
43. Booth, S. C., Meacock, O. J. & Foster, K. R. Cell motility empowers bacterial contact weapons. *ISME J.* **18**, (2024).
44. Gallegos-Monterrosa, R. & Coulthurst, S. J. The ecological impact of a bacterial weapon: microbial interactions and the Type VI secretion system. *FEMS Microbiol. Rev.* **45**, (2021).
45. Ho, B. T., Dong, T. G. & Mekalanos, J. J. A view to a kill: the bacterial type VI secretion system. *Cell Host Microbe* **15**, 9–21 (2014).
46. Salomon, D., Gonzalez, H., Updegraff, B. L. & Orth, K. *Vibrio parahaemolyticus* type VI secretion system 1 is activated in marine conditions to target bacteria, and is differentially regulated from system 2. *PLoS ONE* **8**, e61086 (2013).
47. Si, M. *et al.* Manganese scavenging and oxidative stress response mediated by type VI secretion system in *Burkholderia thailandensis. Proc Natl Acad Sci USA* **114**, E2233–E2242 (2017).
48. Wang, T. *et al.* Type VI secretion system transports Zn^{2+} to combat multiple stresses and host immunity.

PLoS Pathog. **11**, e1005020 (2015).
49. Si, M. *et al.* The Type VI Secretion System Engages a Redox-Regulated Dual-Functional Heme Transporter for Zinc Acquisition. *Cell Rep.* **20**, 949–959 (2017).
50. Barbosa, V. A. A. & Lery, L. M. S. Insights into *Klebsiella pneumoniae* Type VI secretion system transcriptional regulation. *BMC Genomics* **20**, 506 (2019).
51. Hehemann, J.-H. *et al.* Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat. Commun.* **7**, 12860 (2016).
52. Peoples, N., Burns, M. D., Mihalitsis, M. & Wainwright, P. C. Evolutionary lability of a key innovation spurs rapid diversification. *Nature* (2025) doi:10.1038/s41586-025-08612-z.
53. Mayr, E. *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. 356–366 (1942).
54. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. in *Proceedings of the Sixth International Congress of Genetics* (1932).
55. Rosser, N. *et al.* Hybrid speciation driven by multilocus introgression of ecological traits. *Nature* **628**, 811–817 (2024).
56. Sultana, S. *et al.* Vitamin B12 release through bacteriophage-mediated cell lysis of the marine bacterium *Sulfitobacter* sp. M39. *ISME Commun.* **5**, ycaf136 (2025).
57. Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The evolution and ecology of bacterial warfare. *Curr. Biol.* **29**, R521–R537 (2019).
58. Lien, Y.-W. *et al.* Mechanism of bacterial predation via ixotrophy. *Science* **386**, eadp0614 (2024).
59. Nadell, C. D. & Marx, C. J. Bacteria poison and eat their neighbors. *Science* **388**, 1134–1135 (2025).
60. Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J. & Polz, M. F. A reverse ecology approach based on a biological definition of microbial populations. *Cell* **178**, 820-834.e14 (2019).
61. Zhao, W., Caro, F., Robins, W. & Mekalanos, J. J. Antagonism toward the intestinal microbiota and its effect on *Vibrio cholerae* virulence. *Science* **359**, 210–213 (2018).
62. Otto, S. B., Servajean, R., Lemopoulos, A., Bitbol, A.-F. & Blokesch, M. Interactions between pili affect the outcome of bacterial competition driven by the type VI secretion system. *Curr. Biol.* **34**, 2403-2417.e9 (2024).
63. GitHub - rrwick/Filtlong: quality filtering tool for long reads. https://github.com/rrwick/Filtlong.
64. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
65. Wick, R. R. *et al.* Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* **22**, 266 (2021).
66. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
67. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
68. Page, A. J. *et al.* SNP-sites
: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2**, e000056 (2016).
69. Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**, 4310–4312 (2018).
70. Weir, B. S. & Cockerham, C. C. Estimating f-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
71. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
72. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
73. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
74. Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
75. Freese, N. H., Norris, D. C. & Loraine, A. E. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**, 2089–2095 (2016).
76. Chimalapati, S., Lafrance, A. E., Chen, L. & Orth, K. *Vibrio parahaemolyticus:* Basic Techniques for Growth, Genetic Manipulation, and Analysis of Virulence Factors. *Curr. Protoc. Microbiol.* **59**, e131 (2020).

77. Mou, X., Spinard, E. J., Driscoll, M. V., Zhao, W. & Nelson, D. R. H-NS is a negative regulator of the two hemolysin/cytotoxin gene clusters in *Vibrio anguillarum*. *Infect. Immun.* **81**, 3566–3576 (2013).
78. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4**, e6441 (2009).
79. Corcoran, C. P. *et al.* Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.* **84**, 428–445 (2012).
80. Morales, V. M., Bäckman, A. & Bagdasarian, M. A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene* **97**, 39–47 (1991).
81. Sharma, C. M., Darfeuille, F., Plantinga, T. H. & Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* **21**, 2804–2817 (2007).
82. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, e107 (2023).
83. Hoffmann, S. *et al.* A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.* **15**, R34 (2014).
84. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
86. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
87. Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
88. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
89. Galperin, M. Y. *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2021).
90. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
91. Danisch, S. & Krumbiegel, J. Makie.jl: Flexible high-performance data visualization for Julia. *JOSS* **6**, 3349 (2021).
92. Wang, Z. *et al.* Gene essentiality profiling reveals a novel determinant of stresses preventing protein aggregation in *Salmonella*. *Emerg. Microbes Infect.* **11**, 1554–1571 (2022).
93. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279–289 (2009).
94. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
95. Barquist, L. *et al.* The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109–1111 (2016).