The Siderophore Metabolome of

*Azotobacter vinelandii*

Oliver Baars\(^a\), Xinning Zhang\(^a\), François M. M. Morel\(^a\)#, Mohammad R. Seyedsayamdost\(^b\)#

Department of Geosciences\(^a\) and Department of Chemistry\(^b\), Princeton University, Princeton, New Jersey, USA

# Address correspondence to: Mohammad R. Seyedsayamdost (mrseyed@princeton.edu) or François M. M. Morel (morel@princeton.edu)

Running Head: Siderophore function and diversity in *A. vinelandii*
ABSTRACT

In this study, we perform a detailed characterization of the siderophore metabolome, or ‘chelome’, of the agriculturally important and widely-studied model organism *Azotobacter vinelandii*. Using a new high-resolution LC-MS approach, we find over 35 metal-binding secondary metabolites, indicative of a vast chelome in *A. vinelandii*. These include vibrioferrin, a siderophore previously only observed in marine bacteria. Quantitative analyses of siderophore production during diazotrophic growth with different sources and availabilities of Fe show that, under all tested conditions, vibrioferrin is present at the highest concentration of all siderophores and suggest new roles for vibrioferrin in the soil environment. Bioinformatic searches confirm the capacity for vibrioferrin production in *Azotobacter* spp. and other bacteria spanning multiple phyla, habitats, and lifestyles. Moreover, our studies reveal a large number of previously unreported derivatives of all known *A. vinelandii* siderophores and rationalize their origins based on genomic analyses, with implication for siderophore diversity and evolution. Together, these insights provide clues as to why *A. vinelandii* harbors multiple siderophore biosynthetic gene clusters. Coupled with the growing evidence for alternative functions of siderophores, the vast chelome in *A. vinelandii* may be explained by multiple, disparate evolutionary pressures that act on siderophore production.
Azotobacter vinelandii is a widespread nitrogen-fixing soil bacterium belonging to the γ-proteobacteria. It is an established, genetically tractable model organism for studies of nitrogen fixation and siderophore production (1). Siderophores are Fe chelating molecules that change the speciation of Fe in the extracellular medium by outcompeting other natural ligands (2). Uptake of the resulting Fe-siderophore complex via membrane-bound receptors allows A. vinelandii to gain access to otherwise sparingly-soluble Fe (3-5). The Fe-siderophore complexes may be unavailable to competing organisms and thus may exhibit growth-inhibitory or anti-phytopathogenic activities (6, 7). Several studies have shown that the siderophores of A. vinelandii can also bind metals other than Fe to enable uptake of additional metals required in nitrogenases (Mo, V) (4, 8), or to sequester toxic heavy metals (e.g. W, Zn) (9-11). The siderophores excreted by A. vinelandii have also been found to support the growth of some freshwater algae in co-culture by providing a significant source of nitrogen to these organisms (12).

The known siderophores of A. vinelandii include the fluorescent compounds azotobactin D and azotobactin δ (13, 14), and the catechol siderophores, azotochelin (15), aminochelin (16), and protochelin (3). These five siderophores have been discovered and characterized over a span of about 30 years using primarily chemical assays (17), which allow the analysis of only one or few siderophores at the same time due to limited sensitivity and separation power. Thus, it is possible that A. vinelandii produces other, yet unidentified siderophores.

A recent development in the discovery of siderophores is the use of high-resolution liquid-chromatography electrospray ionization mass spectrometry (HR-LC-MS) methods that exploit the characteristic $^{54}$Fe-$^{56}$Fe isotope pattern associated with organic Fe chelates (18-20). Data mining techniques are available for filtering the relevant Fe isotope patterns associated with Fe
complexes even at low abundances and in highly complex matrixes, as well as for detecting the corresponding apo-siderophores (18). Characterization of the species thus discovered can then be achieved by analysis of tandem MS spectra and additional spectroscopic data (e.g. UV-vis, NMR).

Parallel to Fe detection approaches, our understanding of siderophore biosynthesis has increased immensely over the last decade to the extent that bioinformatic mining of genomes can reveal gene clusters responsible for siderophore production, although the exact chemical structure of the final products is often difficult to predict (21-23). Non-ribosomal peptide synthetase (NRPS) genes involved in the production of the known azotobactin and catechol siderophores have been identified in *A. vinelandii* (24, 25).

In this study, we have combined bioinformatic analyses with untargeted HR-LC-MS to discover siderophores and their biosynthetic gene clusters in *A. vinelandii*. The results provide a number of new insights: the α-hydroxycarboxylate siderophore vibrioferrin is observed for the first time in a terrestrial organism and detected at higher concentrations than any of the other siderophores. In addition, a large number of new derivatives of vibrioferrin, azotobactin, and the catechol siderophores, some of which we assign by MS/MS spectral networking methods, have been identified. Finally, functional studies provide insights into possible roles of these siderophores, which begin to explain why *A. vinelandii* carries multiple siderophore biosynthetic gene clusters.

**MATERIALS AND METHODS**

**Bacterial cultures.** Batch cultures of wild type *A. vinelandii* strain CA (also known as strain OP and ATCC 13705) were grown aerobically in a modified Burk’s medium
[(glucose]=10 g L⁻¹; [mannitol] = 10 g L⁻¹; [KH₂PO₄]=5 mM; [K₂HPO₄]=2.3 mM; [CaCl₂] = 0.68 mM; [MgSO₄]=0.41 mM; pH=6.7) under diazotrophic, Fe-limiting conditions by shaking at room temperature (8). Fe bioavailability was controlled by the addition of 100 μM EDTA and 0.1 μM FeCl₃ in HR-LC-MS experiments. Mo concentration ([Na₂MoO₄]=1 μM) was higher than required for optimal growth. Other trace metals were supplemented at optimal concentrations ([CuCl₂]=10⁻⁸ M; [MnCl₂]=2.25x10⁻⁷ M; [CoCl₂]=2.43x10⁻⁸ M; [ZnSO₄]=5.3x10⁻⁸ M) (8). To study the effect of Fe sources, Fe was added as 1) 100 μM EDTA and 0.1 μM FeCl₃; 2) 100 μM EDTA and 5 μM FeCl₃; 3) hematite; and 4) freshly precipitated Fe oxides. Other media components remained the same as previously described. Bacterial growth was monitored at an OD of 620 nm (OD₆₂₀).

**Genome mining.** The genome of *A. vinelandii* strain CA (Genbank accession CP005094) was analyzed for siderophores using the secondary metabolite genome mining software, AntiSmash (26), the gene annotation software, RAST (27), and targeted BLAST homology analyses. The results were compared to previously published studies on siderophore synthesis in *A. vinelandii* (24, 25). A concatenated amino acid sequence corresponding to the entire PvsABCDE cluster in *A. vinelandii* was used for vibrioferrin gene discovery in other publically available genomes based on a tBLASTn search of the NCBI database.

**HR-LC-MS and siderophore metabolomic analyses: sample preparation.** For untargeted siderophore profiling, stationary phase cultures were first collected by centrifugation. The supernatant was filtered, first with a 0.22 μm filter, then with a 3 kDa-cutoff Amicon Ultra ultrafiltration device. Trichloroacetic acid (TCA) was added to a final concentration of 0.03% (v/v) before solid-phase extraction (SPE). After loading, the column (Oasis HLB, 200mg, Waters) was washed with TCA (0.03% in water), then formic acid (FA, 0.03% in water),
followed by elution with 50% and 100% methanol in water. For further analyses, both extract fractions were combined. Sterile culture medium was extracted in the same way and used as a blank. Biological control samples extracted under oxygen free conditions in an anaerobic glove box (Coy chamber) were harvested during early growth (OD620=0.10). Oxygen concentrations in the culture medium were near the detection limit (0.01-0.02 mg L\(^{-1}\)) with a Hach HQ40d oxygen electrode before extraction. Solvents used for these anaerobic extractions were degassed to reach oxygen levels below the detection limit. A control sample was extracted in the same way as described above, another control sample was extracted without the TFA or FA additions, at neutral pH. Concentrated methanolic extracts were dried in a SpeedVac (ThermoFisher) and reconstituted with aqueous mobile phase buffer prior to LC-MS analyses.

**HR-LC-MS measurement.** HR-LC-MS analyses were performed on a high mass accuracy and resolution, reversed phase HPLC-MS platform, using a C\(_{18}\) column (ACE 3 C18-AR, 1mm x 10cm, MAC MOD) coupled to an LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher). Injected samples (5 µL) were separated (1 hr) under a gradient of solutions A and B (solution A: water + 0.1% FA + 0.1% acetic acid; solution B: acetonitrile + 0.1% FA + 0.1% acetic acid; gradient 0-100% B, flow rate 50 µL/min). The control samples were additionally measured using an ammonium acetate (NH\(_4\)OAc) mobile phase buffer (pH=5.0) with the same gradient (solution A: 5 mM NH\(_4\)OAc in water, solution B: 5 mM NH\(_4\)OAc in acetonitrile). To resolve some co-eluting compounds, the control samples were run with a nano-flow capillary ultra-high performance LC system (Nano Ultra 2D Plus, Eksigent, Dublin, CA) coupled to the same LTQ-Orbitrap XL mass spectrometer. These control samples (7 µL) were loaded for a period of 30 min, followed by 1 h of separation over the analytical capillary column (75 um x ca. 25 cm capillary packed with Magic AQ 3um C18 resin). Full-scan mass spectra were acquired in
positive-ion mode ($m/z = 153–1500$) with an experimental resolving power of $R=60000$ ($m/z=400$). MS/MS spectra were simultaneously acquired using CID in the Orbitrap using a parent ion intensity threshold $>10000$ NL and targeting the three most abundant species in the full-scan spectrum or selectively only pre-defined species on a parent ion list.

**Data processing and analysis.** For a schematic representation of the LC-MS analysis workflow, see Fig. 1. The HR-LC-MS dataset was filtered for the characteristic $^{54}\text{Fe}^{56}\text{Fe}$ isotope pattern that is associated with Fe complexes using the software ChelomEx (18). The desired Fe complexes (e.g. $[\text{M}-2\text{H}^{+}+\text{Fe}^{3+}]^{+}$ or $[\text{M}-\text{H}^{+}+\text{Fe}^{2+}]^{+}$), co-eluted with species that had $m/z$ values corresponding to the free siderophore ($\text{MH}^{+}$) and were present at about 100-1000x higher intensities than the Fe complex. To achieve high sensitivity in the detection of Fe chelates, a filter was applied that required the presence of at least 1 matched isotope pattern ($\Delta m/z^{54}\text{Fe}^{56}\text{Fe} = -1.9953 \pm (0.0015 + 2\text{ppm})$, relative intensity $^{54}\text{Fe}^{56}\text{Fe} = 0.064 (0.03-0.09)$) around the apex of the peak and co-elution of the Fe complex with the free ligand (for Fe(III): $\Delta m/z=-53.91928 \pm(0.0015 +2\text{ppm})$; for Fe(II): $\Delta m/z=-52.91145 \pm(0.0015 +2\text{ppm})$). To collect MS/MS spectra of possible siderophores, a parent ion list was generated using an extended list of possible siderophores by including also species that may not show an identified Fe isotope pattern because the $^{54}\text{Fe}$ intensity was below the detection limit, but still showed co-elution with species corresponding to the possible free ligand, whereby the intensity of the free ligand had to exceed that of the Fe complex by $>5x$. The results were examined manually and the free ligands of all species discovered as possible siderophores were included in the parent ion list for high resolution MS/MS data acquisition in replicate runs. In further analyses, only those species were considered, that were present in three replicate runs and also in biological replicate controls, extracted under exclusion of oxygen in the glovebox, but not in the blank media extract. Most
known siderophores fall into a mass range between 400-1500 (23). While we observed several
unknown Fe chelators with masses <400, they formed Fe complexes with two or more ligands
and included also apparent mixed ligand complexes. The binding of these species is less specific
than those of true siderophores and may include, for example fatty acids or amino acids. For this
reason, we excluded species with \( m/z < 400 \) from further analysis.

**MS/MS molecular networks.** MS/MS spectra with the same parent ion mass (±5 ppm or
0.0025 amu) were averaged, de-noised, and the \(^{13}\text{C} \) isotopes were removed. All species selected
for calculation of the network had significantly different retention times. Co-eluting species were
manually examined and possible adducts (e.g. Na or K adducts), dimers, or apparent in-source
fragmentation products were removed. The network was then generated by modification of
methods previously described by Dorrestein and coworkers to adapt to high-resolution MS/MS
spectra (28-30). Briefly, cosine scores were calculated for pairwise aligned MS/MS spectra
reflecting similarity whereby 1 and 0 indicate identical spectra and no similarity, respectively.
Two MS/MS peaks were matched if they had the same high-resolution mass (±0.005 amu) or if
their masses differ exactly by the mass difference of the two parent ions (±0.005 amu). Only
fragments with an \( m/z \) difference of >=50 to the parent ion were used so as to exclude unspecific
losses (e.g. \( \text{H}_2\text{O}, \text{NH}_3 \)) while including, for example, a loss of the lightest amino acid alanine. A
database of known siderophore structures assembled previously (23) was used to match the
masses of molecules and their MS/MS fragments (±5 ppm or 0.005 amu) and to reconstruct
MS/MS spectra.

**Targeted siderophore quantification during growth.**
Targeted quantification of siderophore was performed on a single quadrupole LC-MS system
(Agilent 6120), equipped with a UV-vis spectrometer, in single ion monitoring (SIM) mode.
Sample aliquots of 1 mL were taken throughout the growth, sterile-filtered through 0.2 μm syringe filters, and stored at -20°C until analysis. Prior to analysis the samples were acidified with 0.1% acetic acid and 0.1% FA. Without further purification (no solid phase extraction), 100 μL sample aliquots were injected onto a C18 column (Agilent Eclipse Plus C18 3.5 μm, 4.6x100 mm) equipped with a matching guard column. The separation proceeded with the same mobile phase system as described in the text (solution A: water + 0.1% FA + 0.1% acetic acid; solution B: acetonitrile + 0.1% FA + 0.1% acetic acid) over 30 min, at a flow rate of 0.8 mL/min. Using a 6 port valve, the column outflow was diverted to waste for the first 5.25 min ensuring that the sample was completely desalted before introduction into the mass spectrometer. For quantification, LC-MS and UV-vis peak areas were determined using MassHunter software (Agilent). Relative elution times of the peaks on this system were matched to the elution times for the siderophores determined on the HR-LC-MS system. Peak areas were converted to concentrations by calibration with isolated standards of vibrioferrin, DHBA, azotochelin, protochelin, and azotobactin δ. The concentrations of minor derivatives were estimated using the LC-MS response determined for the structurally closely related major siderophores. Seven technical replicates of a spent medium ‘standard’ collected in the stationary phase from a culture grown under the same conditions as the samples used for HR-LC-MS analysis (100 μM EDTA and 0.1 μM FeCl3) showed relative standard deviations of <3.5% for the vibrioferrins and the major catechol siderophores. The remaining siderophores were measured with slightly larger standard deviations (<10% for siderophore concentrations above 0.5 μM and <20% for lower concentrations).

**Siderophore isolation and quantification.** Isolation of siderophores was achieved by filtration and solid-phase extraction (Oasis MAX or Oasis HLB) of culture media followed by
HPLC purification with a C18 column. The pooled fractions were lyophilized and reconstituted with D$_2$O (vibrio ferrin, aminochelin) or d$_4$-MeOH to obtain $^1$H-NMR and COSY spectra (Bruker Avance III 500MHz). Quantification of the isolated siderophore standards was performed by $^1$H-NMR with internal standard addition of sodium benzoate for vibrioferrin or by UV-vis using reported extinction coefficients for acidified solutions of DHBA, aminochelin, azotochelin, protochelin and azotobactin (5, 31).

RESULTS

Bioinformatic analyses of the *A. vinelandii* siderophore metabolome. We began by mining the genome of *A. vinelandii* strain CA (also known as OP, accession CP005094.1) for siderophore biosynthetic genes and found a total of 9 NRPS genes and 2 NRPS-independent siderophore synthetase genes arranged across 5 clusters (Fig. 2, Table S1). Two of these clusters (AvCA_21160 - AvCA21230 and AvCA_25530 – AvCA_25660) contain NRPS genes previously shown to be necessary for the production of catechols and azotobactin (24). Closer analysis of AvCA_09300 - AvCA_09360, one of the three new gene clusters, indicated that it possibly encodes the synthesis of a vibrioferrin-like compound, an $\alpha$-hydroxycarboxylate siderophore characteristic of marine bacteria such as *Vibrio parahaemolyticus* (32). Bioinformatics-based predictions of the specific siderophores produced by AvCA_09680 and AvCA_09690 are less clear; however, homology searches and the domain structure of these NRPS genes hint at an involvement in the later stages of protochelin biosynthesis (see below). The product of the last gene cluster, containing an MbtH-encoding gene (AvCA_50380) and a NRPS possessing only a single domain (AvCA_50370), is not known.
Discovery of unknown siderophores from *A. vinelandii*. To examine the product(s) of the *A. vinelandii* siderophore metabolome, we cultured *A. vinelandii* strain CA under diazotrophic and Fe-limited conditions, which have previously been shown to stimulate siderophore production (8). Initial mining of the HPLC-MS data for the $^{54}\text{Fe}^{56}\text{Fe}$ isotope pattern associated with Fe(III) or Fe(II) siderophore complexes ($[\text{M-2H}^++\text{Fe}^{3+}]^+$ or $[\text{M-H}^++\text{Fe}^{2+}]^+$) and their related free ligands ($\text{MH}^+$) according to the scheme described in Fig. 1 revealed all previously known siderophores (green in Fig. 3), but also a large number of other possible Fe-chelating agents. The Fe-complexes of all species co-eluted with their corresponding, apo-ligands, which were significantly more abundant than the Fe complexes as expected from the acidic (pH~2.5) and low Fe conditions used for chromatography.

Among the most abundant Fe chelators were species that matched the $m/z$ values of the known siderophores produced by *A. vinelandii* CA, i.e. the fluorescent siderophores azotobactin D and azotobactin δ, as well as the catechol siderophores, aminochelin, azotochelin, and protochelin (green in Fig. 3). The identity of these known siderophores was further established by their MS/MS fragmentation patterns (Fig. S1). Moreover, we found a large number of high-abundance Fe chelators with as of yet unassigned structures. Together, over 35 siderophores were reproducibly found in three biological replicates. These include hydrophilic siderophores, previously not reported in *A. vinelandii* (red, Fig. 2), as well as new derivatives of old siderophores (orange, Fig. 3).

Siderophore molecular networks: identification of vibrioferrins, azotobactins and catechol siderophores. To obtain an overview of the structural similarity and diversity among *A. vinelandii* siderophores, we collected HR-MS/MS data for each compound and created a spectral network, an approach recently pioneered by the Dorrestein group (29). In this network, each
node represents an individual siderophore (adducts, dimers, etc. were manually removed), while
lines connecting the nodes represent commonalities in MS/MS fragmentation (Fig. 4A). The
network revealed three groups of compounds structurally related to azotobactins, catechol
siderophores, and vibrioferrin, a siderophore previously not observed in A. vinelandii.

**Vibrioferin cluster.** The most prominent peak in this group of compounds was a
hydrophilic Fe-chelator that had the same mass as the siderophore vibrioferrin (33), which has
previously only been detected in marine bacteria (m/z=435.125, Fig. 4B). Purification of this
siderophore and subsequent analysis by MS/MS and NMR (1H & COSY) confirmed its identity
(Fig. S2). This assignment is consistent with bioinformatic analysis of the A. vinelandii genome,
which reveals a conserved *pvs* gene cluster (AvCA_09300-AvCA_09360), responsible for
vibrioferrin biosynthesis in *V. parahaemolyticus* (34) (Figs. 2, S2). Vibrioferrin is likely
assembled from citrate, Ala, ethanolamine (derived from Ser), and α-ketoglutarate (35). We
identified two new derivatives, one where the Ala precursor was replaced with Ser (vibrioferrin
B), and another bearing a methyl ester within the citrate substructure (vibrioferrin C). The
structural assignment of these analogs is based on HR-MS/MS data and on isotopic feeding
experiments with 3-2H3-Ala, which resulted in a 3 Da mass shift with vibrioferrins A and C, but
not B, consistent with the structures in Fig. 4B (Fig. S3). These new derivatives are consistent
with precursor flexibility and post-synthetic tailoring in the vibrioferrin biosynthetic pathway.

**Catechol cluster.** In addition to the known catechol siderophores, the HR-LC-MS data
revealed a number of derivatives that differed by simple chemical modifications (+/-O; +/-2H;
CH2 groups). Larger differences corresponded to addition or loss of a dihydroxybenzoyl group
(C6H4O3, Δm=136.013) or an aminochelin group (C11H14O3N2, Δm= 222.100), the building
blocks of azotochelin and protochelin. Based on MS/MS data, structures could be assigned to
several noteworthy derivatives of protochelin, which we denote protochelin B-G (Figs. 4C, S4-S6). For protochelin B, we detected a mass difference of \( \Delta m/z = 14.016 \) relative to protochelin A for both the parent ion and some of its fragments, indicative of a loss of CH\(_2\). Manual analysis of the MS/MS data revealed that the modification was located in the CH\(_2\) chain of the aminochelin residue in protochelin as indicated by red circles in Fig. 4C (see also Fig. S4A). Protochelin C-E, with \( m/z = 489.234 \), were assigned as analogs without a dihydroxybenzoyl group (DHB, Fig. S4B). Three chromatographic peaks, each associated with different MS/MS fragmentation patterns, corresponded to the three structural isomers in which the DHB is missing in each of the three possible positions of protochelin. One of these analogs, protochelin C, is shown in Fig. 4C.

MS data for protochelin F showed that it lacked two hydrogens relative to protochelin A. UV-vis spectra showed peak absorbances at \( \lambda = 330 \) nm in protochelin and in protochelin F indicating that the structural change in the new analogue was not associated with an extended \( \pi \)-electron system. Instead, the aromatic substitution pattern inferred from \(^1\)H-NMR spectra showed that the new compound was characterized by cross-linked catechol rings in agreement with the MS/MS spectra (Figs. S5, S6). The structure for the most prominent peak, protochelin F, is shown (Fig. 4C). A similar compound, protochelin G, with a different aromatic substitution pattern was also isolated (Fig. S6).

**Azotobactin cluster.** In the azotobactin network, several prominent siderophores had a larger mass than the known azotobactins (azotobactin \( \delta \) and azotobactin D). The structure of azotobactin \( \delta \) (\( m/z \) for \([M+2H]^{2+} = 697.261, z=2\)) is characterized by a dihydroxyquinoline chromophore and a peptide chain with a homoserine-lactone at the terminus of the peptide chain, which are marked in orange and green, respectively in Fig. 4D. Azotobactin D (\([M+2H]^{2+} = 706.266\)) has the same structure as azotobactin \( \delta \) with a terminal homoserine instead...
of the homoserine-lactone. The masses of the newly identified azotobactin derivatives match the masses of putative azotobactin precursors, which are characterized by modifications of the chromophore with a glutamic acid side chain. These structures have been suggested to occur in the following oxidative cascade in pseudomonads: ferribactin $\rightarrow$ hydroxyl-ferribactin $\rightarrow$ dihydroxyverdine $\rightarrow$ pyoverdine $\rightarrow$ azotobactin (36-38) (Fig. 4D). Changes in the putative chromophore structures were also in agreement with expected shifts in the UV-vis spectra. The hydroxyl-ferribactin $\delta$ ([M+2H]$^{2+}$ = 750.807) was the most abundant azotobactin derivative in this study. All azotobactin derivatives in Fig. 4D had the same peptide chain as the previously known azotobactins and a structural change in the chromophore with an additional glutamic acid side chain. An exception is a compound with an additional CH$_2$ group ([M+2H]$^{2+}$=713.274), which we denote azotobactin D2. MS/MS spectra indicated that the additional CH$_2$ was located at the homoserine end of azotobactin D2, likely representing a methyl ester or methyl ether derivative of homoserine (green in Fig. 4D).

Pyoverdines, siderophores produced by pseudomonads, chemically closely related to azotobactins, show a remarkable diversity in their peptide structures between species and strains (e.g. *P. aeruginosa*, (39)) that is encoded in the NRPS genes. Nonetheless, in the three closely related *A. vinelandii* strains that have been sequenced (DJ, CA, CA6 (1, 40)) the NRPS genes are 100% identical.

**Siderophore production during growth.** Why does *A. vinelandii* simultaneously produce three structurally distinct groups of siderophores? We sought to answer this question by studying the effect of different Fe sources on siderophore production. *A. vinelandii* was cultured under diazotrophic conditions with four different sources of Fe: 1) 0.1 µM Fe + 100 µM EDTA, i.e. the same conditions used for HR-LC-MS siderophore discovery; 2) 5 µM Fe + 100 µM
EDTA; 3) hematite; and 4) freshly precipitated amorphous Fe oxide. We quantified siderophores by direct injection on a quadrupole LC-MS without prior solid phase extraction thus avoiding possible analytical errors from low analyte recoveries or pre-concentration artifacts.

Under conditions 2 and 4 the cells grew rapidly and reached high optical densities (OD) after three days, while slow growth and low maximum OD indicated severe Fe-limitation under conditions 1 and 3 (Fig. 5). Vibrio ferrin A was the major siderophore under all tested growth conditions, present at 4-14x the concentration of any of the other siderophores (Fig. 5).

Vibrio ferrin A was detected at particularly high concentrations, in the late exponential growth phase under condition 2 (high Fe + EDTA), reaching concentrations up to 360 µM. Vibrio ferrin derivatives were present at ~10x lower concentrations (vibrio ferrin B: 3-11 µM; vibrio ferrin C: 0-4 µM). The abundant production of vibrio ferrins under condition 2 (high Fe + EDTA) was followed by a rapid production of aminochelin and azotochelin which reached final concentrations of 60 and 75 µM respectively. In contrast, when grown with highly available amorphous Fe oxide, aminochelin and azotochelin concentrations remained very low (<1.5 µM) while intermediate concentrations (7-10 µM) were reached under conditions of severe Fe limitation (added 0.1 µM Fe + 100 µM EDTA or added hematite). Interestingly, the production of protochelin was remarkably insensitive to the Fe source, possibly due to its role as a metallophore for Mo (8), which was added at a concentration close to the observed maximum protochelin concentrations in all treatments ([Mo]=1 µM). The most abundant protochelin derivative, protochelin D, was found at similar concentrations as protochelin A (~1 µM) but it was not formed under condition 4. Azotobactins and related compounds were only observed under severe Fe limitation (conditions 1 and 3) in agreement with previous observations (4, 41).

Azotobactins D and δ reached concentrations of <1 µM. The newly identified azotobactin-related
siderophores hydroxyl-ferribactin δ and hydroxyl-ferribactin D were present at concentrations of up to 2 µM. The hydroxyl-ferribactins were detected at higher concentrations than the azotobactins under condition 3 (hematite) but lower concentrations under condition 1 (low Fe + EDTA).

The relative timing of production of each siderophore was roughly similar under all conditions: (1) protochelin synthesis during the initial lag and early growth phase, (2) massive vibrioferrin production after an initial fast growth phase and (3) production of, aminochelin, azotochelin, and the known azotobactins starting late in the initial rapid growth phase and continuing in the later growth phase. The major azotobactin derivatives, hydroxyl-ferribactin δ and hydroxyl-ferribactin D increased earlier than the known azotobactins. Production of vibrioferrins B and C followed that of vibrioferrin A, while the protochelin derivatives peaked later than protochelin A. These results provide some clues regarding the role of each siderophore and have implications for the ability of bacteria to synthesize numerous siderophores, as discussed below.

DISCUSSION

Complete descriptions of siderophore metabolomes are a necessary precondition to address successfully the important and outstanding question of why microorganisms produce simultaneously a large set of distinct siderophores, and why this set is different in different species. By exploiting state-of-the-art HR-LC-MS analysis combined with new data processing techniques, this study provides the most complete analysis to date of siderophores produced by any micro-organism. Aside from the previously known azotobactins and catechol siderophores, we have identified the presence of additional, abundant, and previously unreported siderophores.
produced by *A. vinelandii*, including the hydrophilic vibrioferrin and its derivatives as well as analogs of azotobactin, azotochelin, protochelin, and aminochelin (Table S2). These findings raise questions regarding the biosynthetic origins and possible functions of the new siderophores.

**Vibrioferrin production by *A. vinelandii* and other bacteria.** Vibrioferrin is the siderophore produced at the highest concentration under all tested conditions (diazotrophic growth with varying sources and availabilities of Fe), even though it has not been detected previously in the *A. vinelandii* growth medium (Fig. 5). One possible explanation is that vibrioferrin is more hydrophilic than the previously known *A. vinelandii* siderophores and might not be retained by some reversed phase extraction protocols. Vibrioferrin production was first observed in *V. parahaemolyticus* and since then has been found in several marine bacteria including *Marinobacter* symbionts of dinoflagellates (42, 43). Our study provides the first example of vibrioferrin production outside of the marine environment. Genomic analysis for *pvs* genes, previously shown to be responsible for vibrioferrin biosynthesis (32), reveals striking conservation of the *pvsABCDE* gene cluster in bacteria from multiple phyla, environments, and life styles (Fig. 6). These genes are primarily found in γ-proteobacteria, including several non-marine organisms that are free-living or pathogenic. Thus, the potential for vibrioferrin production is more widespread than previously recognized. Additionally, the presence of *pvs* genes on a plasmid of *Ralstonia eutropha*, a β-proteobacterium, suggests that bacterial conjugation may be a mechanism for horizontal transfer of vibrioferrin genes.

In *A. vinelandii*, vibrioferrin production was particularly pronounced during the late exponential growth phase in media with high initial added Fe (5 μM) and an excess of EDTA (100 μM) (condition 2, Fig. 5). Under the same condition, we also observed the highest concentrations of aminochelin and azotochelin. This strong co-production of vibrioferrin,
aminochelin, and azotochelin during the late exponential phase likely reflected the low concentrations of available Fe related to the slow dissociation of the Fe-EDTA complex, which cannot keep up with the demands of the multiplying bacterium (44, 45). Our finding of abundant vibrioferrin production outside of a marine environment and without light indicates that Fe is made available to bacteria via mechanisms other than the previously suggested photoreduction of Fe-vibrioferrin (42, 43). Vibrioferrin is hydrophilic and a weak Fe chelator (pFe=-log $[Fe^{3+}] = 18.4$ with $[vibrioferrin] = 10^{-5}$ M, $[Fe] = 10^{-6}$ M, pH=7.4) (43). The strong co-production of the weak, hydrophilic, and concentrated vibrioferrin with the stronger, more hydrophobic, and less concentrated azotochelin (pFe=23.1 (31)) may be required in a ‘bucket brigade’ mechanism that involves binding of Fe by vibrioferrin in the bulk medium and exchange with a hydrophobic siderophore that delivers Fe to the cell (46, 47). As such, the disparate siderophores would act synergistically in the acquisition of Fe (48). Notably, several bacteria with vibrioferrin biosynthetic genes in Fig. 6 can potentially also produce other, structurally unrelated siderophores, which may be used in a bucket brigade mechanism.

When grown with amorphous Fe oxide (condition 4), *A. vinelandii* grows as fast and reaches the same OD as with high Fe + EDTA (condition 2), but the siderophore pool consists almost exclusively of vibrioferrin and DHBA, with aminochelin and azotochelin at very low levels. Notably, both vibrioferrin (pKₐ values 5.1; 3.6; 2.7 (43)) and DHBA (pKₐ = 2.9) are negatively charged at the experimental pH (6.9), unlike aminochelin (pKₐ values 7.1; 10.2; 12.1 (49)), the other hydrophilic siderophore of *A. vinelandii*, which is positively charged. The negative charge of vibrioferrin and DHBA favors adsorption on positively charged Fe-oxyhydroxides (at pH < 8), which may help in the dissolution of Fe and explain their elevated production compared to all...
other siderophores under condition 4. Thus, it appears that *A. vinelandii* can tailor its siderophore metabolome to existing sources of Fe in the environment.

**Siderophore derivatives: artifacts, spontaneous reactions or biosynthesis?** The LC-MS analysis reveals new structural variations for vibrioferrin and the previously described *A. vinelandii* siderophores (Figs. 3-5). These related species were present at lower concentrations and may principally result from methodological artifacts, spontaneous reactions, or targeted biosynthesis. Methodological artifacts may potentially occur during sample preparation or LC-MS measurements, including (1) inadvertent oxidation, (2) reactions due to acidification, (3) reactions during the reversed phase extraction, and (4) reactions during ionization. To rule out oxidation or acidification, we performed control experiments in the absence of oxygen and at pH=5 instead of 2.5 by omitting sample acidification during solid phase extraction and by using an ammonium acetate buffer (pH~5) during LC-MS. All the derivatives shown in Fig. 4 were observed under anoxic conditions and at the higher pH except for azotobactins and their derivatives, which were not retained on the solid phase column or did not ionize at pH 5. The possibility of new species being caused by reversed phase extraction is ruled out by the observation of many of the same derivatives via direct injection of the filtered medium into the LC-MS. Finally, species caused by reactions during the electrospray ionization process should co-elute in the LC-MS chromatogram. Yet we observed separated chromatographic peaks for the siderophores shown in Fig. 4 (see also Table S2). Thus potential methodological artifacts could clearly not explain the observed derivatives of the major *A. vinelandii* siderophores, and these are generated by spontaneous reactions or in an enzyme-dependent, ‘deliberate’ fashion by the bacteria.
Derivatives of the catechol siderophores: biosynthesis and spontaneous reactions. The group of catechol siderophores is particularly diverse and their biosyntheses have been well-studied. We therefore bioinformatically investigated the production of the catechol siderophores. Analysis of the ent gene cluster, previously shown to be involved in catechol biosynthesis (cluster 1, Fig. 7A, AvCA_21180-AvCA_21230), suggests that additional genes are required for production of aminochelin and protochelin (24, 25). This cluster can account for production of an EntB-bound DHB-thioester (Fig. 7B). Production of aminochelin from this intermediate requires incorporation of butane-1,4-diamine (putrescine), a reaction analogous to that catalyzed by VibH (Genbank acc. WP_001880577.1), an amide synthase from Vibrio cholerae that combines the polyamine norspermidine with DHB-thioesters in vibriobactin biosynthesis (50). A closer examination indeed indicates the presence of a VibH-like enzyme, which has 41% similarity (21% identity) to the V. cholerae VibH (Genbank acc. WP_001880577.1), adjacent to the ent gene cluster (AVCA_21160, Fig. 7A). Thus a model can be proposed for production of aminochelin A (Fig. 7B). In an analogous fashion, aminochelin B would be generated from the same EntB-bound DHB intermediate and propane-1,3-diamine (dashed line 1, Fig. 7B), similar to the biosynthesis of serratiochelins, which also utilizes propane-1,3-diamine (51). Both diamines are well-known and among the dominant forms in bacteria (52, 53). Thus, a single VibH-like homolog may incorporate several polyamines, consistent with previous in vitro studies on VibH (50).

For the bioproduction of the bis-catachol azotochelin, we predict that the large NRPS encoded by entF (AvCA21190) catalyzes condensation of a T-domain-bound Lys with the DHB-thioester (Fig. 7C). Subsequent release from the assembly line via the TE domain would furnish azotochelin. On the other hand, the biosynthesis of the protochelins will likely require concerted...
action of an NRPS from a different genetic locus, possibly one or both NRPSs found in cluster 2 (Fig. 7A). Among these, AvCA_09680 is especially intriguing as it contains the domain architecture C*-A-T, where C* represents a modified C-domain with an HHxxxDA signature motif (rather than the canonical HHxxxDG sequence). A similar ‘unusual’ C-domain has been shown to condense a diffusible $N^\text{1}$-(2,3-dihydroxybenzoyl)norspermidine group with T-domain bound DHB-thioester in the biosynthesis of vibriobactin (54). We therefore propose that this unusual NRPS is involved in the production of protochelin and derivatives B, E, and C (dashed lines 2-4, Fig. 7C). In the case of protochelin B, aminochelin B is utilized as the diffusible substrate, whereas biosyntheses of protochelins A/C and E require aminochelin A and putrescine, respectively. Generation of these kinds of analogs, lacking a DHB moiety, was previously also observed for vibriobactins (55). Thus, biosynthetic origins can be proposed for many of the siderophore analogs using bioinformatic analyses.

Some species we have detected may also be formed by spontaneous post-synthesis reactions. In particular, protochelin F may be derived from such a route for 3 reasons: (1) its structure is characterized by a direct intramolecular cross-link between two catechol rings hindering Fe binding; (2) it has several isomers, such as protochelin G, with the link between the catechol rings at different positions arguing against a defined enzymatic synthesis; and (3) its concentration peaks late during growth when protochelin concentrations decrease (Fig. 5). A possible formation mechanism in aerobic A. vinelandii cultures involves oxidation of catechols to form crosslinked aromatic compounds, perhaps catalyzed by the presence of Fe (56, 57).

**Functions of multiple siderophore derivatives.** In this study we observed a large number of previously unreported derivatives of all the siderophores of A. vinelandii. The production of several related siderophores by a single organism has been observed for ferrioxamines,
agrobactins, desferrichromes, enterobactins, mycobactins and pyoverdines (23). The high sensitivity of our LC-MS approach reveals that even with compounds for which previously only one structure was known, such as for vibrioferrin, azotochelin, or protochelin, there are a large number of less abundant derivatives, vastly expanding on the known variations. Our detailed comparison of metabolomic and genetic data for catechol biosynthesis reveal flexible a substrate range and non-linear siderophore assembly as key mechanisms for generating chemical diversity. These reactions give rise to structural modifications with altered hydrophobicity, binding affinities, and kinetics of interaction with uptake transporters that natural selection may act upon during structurally-mediated evolution of new siderophore functions.

Possible advantages of structural variation include the prevention of binding or uptake of siderophores by competing organisms. For example, it has been shown that Streptomycetes produce additional siderophores when grown in co-culture with competing strains (28). Structural variation of siderophores may also allow the organism to optimize uptake depending on environmental Fe chemistries (58), facilitate uptake of other required metals (e.g. Mo and V), or sequester toxic metals (e.g. W) (10). Siderophores may also act as redox shuttles, even as signaling molecules, and can serve multiple functions simultaneously (59). This leads to a fundamental reason for the vast structural variation: siderophore structures could result from multiple selective pressures that reflect evolutionary arms races including not just competition for low abundance catalytic metals, but also other processes. Future untargeted siderophore metabolomics analyses will further improve our understanding of conditions under which siderophores are produced and how these conditions relate to siderophore structural variations and function.
FUNDING INFORMATION

We thank the National Science Foundation (grant OCE 1315200 to F.M.M.M.), the Pew Biomedical Scholars Program (M.R.S.), and the Grand Challenge Program of the Princeton Environmental Institute (F.M.M.M., O.B. & M.R.S.) for financial support of this work.

REFERENCES


15. Corbin JL, Bulen WA. 1969. Isolation and identification of 2,3-dihydroxybenzoic acid and N2,N6-di([2,3-dihydroxybenzoyl]-L-lysine formed by iron-deficient *Azotobacter vinelandii*. Biochemistry 8:757-762.


For untargeted siderophore profiling (‘chelomics’), sample preparation and measurement on a high resolution LC-MS system was followed by mining of the LC-MS data for characteristic Fe isotope patterns associated with Fe complexes, and the presence of associated Fe-free ligand species using dedicated data processing software (ChelomEx, (18)). MS/MS molecular networks of identified putative free siderophores were created to group structurally related species and assign mass differences between related species to sum formulas differences. Identification of new siderophores was assisted by comparison of molecule and fragment masses to a database of known siderophore structures. Finally, manual reconstruction of MS/MS spectra allowed the assignment of chemical structures to several new siderophore structures. In some cases the structure assignment was informed by additional NMR spectra of isolated compounds.
Quantification of identified siderophores was performed on a single quadrupole LC-MS by direct injection of filtered spent media without prior solid phase extraction (SPE).
Fig. 2: Non-ribosomal peptide synthase (NRPS) and NRPS-independent gene clusters in *A. vinelandii* strain CA (Genbank accession CP005094.1). Genes necessary for production of specific siderophores are indicated by superscripts (\(^1\)catechols (24); \(^2\)catechols (25); \(^3\)azotobactin (25)). Annotated function of genes as well as a more complete list of azotobactin related biosynthetic genes can be found in Table S1.
Fig. 3: Base peak chromatogram (BPC) for the high-resolution LC-MS analysis of the *A. vinelandii* spent medium (black). Overlying the BPC are extracted ion chromatograms of the known siderophores from *A. vinelandii* (green). Siderophores that have not been reported before from *A. vinelandii* include the hydrophilic vibrioferrin (red) and a large number of new siderophores (orange), which were found to be related to the known siderophores produced by the bacterium.
Fig. 4: (A) MS/MS molecular network of siderophores produced by *A. vinelandii*. Each node represents a separate siderophore (adducts, dimers, etc. removed) that was required to be present.
in biological and analytical replicates. The thickness of the edge represents relatedness between
the MS/MS spectra of two species. The known siderophores from \textit{A. vinelandii} are indicated as
black circles, vibrioferrin is indicated as a yellow circle. Three separate clusters can be
recognized that include vibrioferrin, catechol siderophores, and azotobactins. The software
Cytoscape was used for visualization. (B), (C) and (D) show nodes in the three clusters in more
detail. For these networks, nodes were manually arranged and only selected edges are shown for
clarity of presentation. The ring color around the nodes represents the peak area for each species
and the number represents the corresponding \textit{m/z} value (rounded to one digit). The exact mass
difference between two nodes was assigned to chemical sum formulas as indicated. Structures of
three notable new catechol siderophore derivatives shown in (C) are based on reconstruction of
MS/MS spectra and additional spectroscopic data (see text). The table in (D) shows the MS/MS
fragmentation of azotobactin related compounds; \( \lambda_{\text{max}} \) is the absorption maximum in the UV-vis
spectrum. Arrows indicate MS/MS fragments corresponding to the B (arrows to the left) and Y
(arrows to the right) ions. Note, that all siderophore species in the azotobactin cluster were
doubly charged. ChrA = azotobactin chromophore; Febn = ferribactin; OHFebn = hydroxyl-
ferribactin; ChrP = pyoverdine chromophore; 2HChrP = dihydropyoverdine; Ser = serine; Glu =
glutamate; Hse = homoserine; Gly = glycine; OHAsp = hydroxylaspartate; Cit = citrulline;
AcOHOrn = acylhydroxyornithine; Hsl = homoserine lactone; MeHse = methylhomoserine
Fig. 5: Concentration of notable siderophores from *A. vinelandii* under diazotrophic growth with different sources and availability of Fe. Growth was monitored by optical density at 620 nm (OD$_{620}$). (*) and (**) indicate new *A. vinelandii* siderophores identified in this study. (**) represent azotobactin derivatives based on MS/MS fragmentation patterns as shown in Fig. 3. Relative standard deviations were <3.5 % for the vibrioferrins and the major catechol
siderophores based on replicate analyses of a representative spent medium ‘standard’. The remaining siderophores were measured with slightly larger standard deviations (<10% for siderophore concentrations above 0.5 µM and <20% for lower concentrations).
Fig. 6: Occurrence of vibrioferin biosynthetic genes (pvs) in bacteria from diverse phyla, environments, and life-styles. Information on gene loci is displayed below each arrow. Genbank accession numbers for genomes shown above (in order) are BA000032.2, CP000316.1, CP005094.1, CP010415.1, CP003057.1, CP000744.1, CP006664.1, CP000316.1, AY305378.1, CP001114.1, BA000030.3.
Fig. 7: Proposed biosynthesis for *A. vinelandii* siderophores. (A) Mining of the *A. vinelandii* strain CA genome suggests two NRPS clusters are involved in catechol siderophore production. (B) Proposed biosynthesis pathway for mono-catechol siderophores. (C) Proposed biosynthesis for bis and tris-catechol siderophores. Bold text indicates previously characterized catechols. Dashed lines indicate pathways for derivatives identified in this study (see Fig. 4).