BMC Research Notes

DATA NOTE

Open Access

The complete genome sequence of *Hafnia alvei* A23BA; a potential antibiotic-producing rhizobacterium

Opeyemi K. Awolope[1] , Noelle H. O'Driscoll[1] , Alberto Di Salvo[1] and Andrew J. Lamb[2*]

Abstract

Objectives: The urgent need for novel antibiotics cannot be overemphasized. *Hafnia alvei* A23BA was isolated from plant rhizosphere as part of an effort to recover novel antibiotic-producing bacterial strains from soil samples. The genome of the isolate was sequenced to facilitate mining for potential antibiotic-encoding biosynthetic gene clusters and to gain insights into how these gene clusters could be activated.

Data description: Here, we report the complete genome sequence of *H. alvei* A23BA obtained from the hybrid assembly of Illumina HiSeq and GridION reads. The genome, consisting of a circular chromosome and a circular plasmid, is 4.77 Mb in size with a GC content of 48.77%. The assembly is 99.5% complete with genomic features including 4,217 CDSs, 125 RNAs, and 30 pseudogenes. Thiopeptide, beta-lactone, siderophore, and homoserine lactone biosynthetic gene clusters were also identified. Other gene clusters of interest include those associated with bioremediation, biocontrol, and plant growth promotion- all of which are reported for *H. alvei* for the first time. This dataset serves to expedite the exploration of the biosynthetic and metabolic potentials of the species. Furthermore, being the first published genome sequence of a soil isolate, this dataset enriches the comparative genomics study of *H. alvei* strains.

Keywords: *Hafnia alvei*, *H. alvei*, Genome mining, Biosynthetic gene clusters, Antibiotics, Rhizobacteria, Bioremediation, Biocontrol, Plant growth-promoting rhizobacteria

Objective

Bacterial secondary metabolites are invaluable sources of novel bioactive compounds. Many clinically useful antibiotics were derived from the secondary metabolites of soil dwelling bacteria [1]. However, only a small fraction of all known species have had their metabolites exploited in this way [2]. To this end, we sought to isolate novel antibiotic-producing bacterial strains from soil samples collected from the rhizosphere as antibiosis occurs naturally within it [3]. *H. alvei* A23BA was recovered as part of this effort.

Hafnia alvei is a Gram-negative, rod- shaped, facultatively anaerobic psychrotrophic bacterium. It is commonly isolated from clinical materials, gastrointestinal tract of animals, plant surfaces, soil, and water [4]. Some strains are commensals of the gastrointestinal tract while others are opportunistic pathogens implicated in both nosocomial and community-acquired infections [5, 6]. It is almost never associated with antibiotic production except for the antimicrobial activities reported for a strain isolated from the gut of honeybees [7]. Phylogenetic studies of this little-known species have shown its pan-genome to be open and dynamic with each strain possessing sets of unique genes [8]. Unique gene acquisition is mainly by horizontal gene transfer, and it reflects the adaptation of strains to their remarkably diverse natural habitats. Strains show considerable metabolic

*Correspondence: a.lamb@rgu.ac.uk
[2] Graduate School, Robert Gordon University, The Ishbel Gordon Building, Garthdee Road, Aberdeen AB10 7QE, Scotland
Full list of author information is available at the end of the article

pathway diversity and varied biosynthetic potentials because of the open pan-genome making them good mining candidates for novel metabolites.

Consequently, the genome of *H. alvei* A23BA was sequenced to enable mining for potential antibiotic-encoding secondary metabolite biosynthetic gene clusters (smBGCs) that show little or no homology to known smBGCs. Furthermore, assembled genomes of *H. alvei* in public repositories are typically of clinical, human or food isolates, to the best of our knowledge, the complete genome sequence of *H. alvei* A23BA represents the first published complete genome sequence of a soil isolate.

Data description

H. alvei A23BA was recovered from the rhizosphere of a garden plant in Aberdeen, Scotland (57.101 N 2.078 W). It was isolated using an ultra-minimal substrate medium (Data file 1) [9]. Upon isolation and strain purification, isolate was cultivated in nutrient broth (Oxoid, UK) at 37 °C for 24 h. Overnight culture was centrifuged and gDNA was extracted from pellets with the DNeasy® Ultraclean® Microbial Kit for DNA Isolation (Qiagen, UK). Isolate was preliminarily identified by 16S rRNA gene sequence comparison as *H. alvei* with 99% identity score.

Libraries were subsequently prepared from extracted gDNA by MicrobesNG (Birmingham, UK) for whole genome sequencing. For Illumina sequencing, libraries were prepared using the Nextera XT Library Prep Kit (Illumina, USA) and sequenced with the Illumina HiSeq system using a 250 bp paired end protocol. For GridION (Oxford nanopore) sequencing, libraries were prepared with Oxford nanopore SQK-RBK004 kit and/or SQK-LSK109 kit with Native Barcoding EXP-NBD104/114 (ONT, UK) using 400-500 ng HMW DNA. Sequencing was performed on a FLO-MIN106 (R.9.4 or R.9.4.1) flow cell in a GridION (ONT, UK).

Illumina sequencing run produced 4,973,530 short reads that were trimmed and paired using Trimmomatic [10] v0.30 with a sliding window quality cut-off of Q15. Ninety eight percent of reads were retained, and quality was assessed with FastQC [11] v0.11.8. Mean phred score across each base position was assessed with MultiQC [12] and found to be ≥ 28 (Data file 2) [13]. GridION sequencing run produced 18,642 reads with the mean read quality score of 10.5 (data file 3) [14] as assessed with NanoStat [15]. Paired short reads and long reads from GridION sequencing were assembled with Unicycler [16] v0.4.8.0. Assembly quality was assessed with QUAST [17] v5.0.2- two contigs (one chromosome and one plasmid) were identified with a total length of 4,772,047 bp, N50 value of 4,687,005 bp and #N's per 100 kbp value of 0 (data file 4) [18]. Assembly completeness was assessed with BUSCO [19] v3.0.2 and found to be 99.5% (data file 5) [20]. Identity was confirmed as *H. alvei* by ANI analysis using the FastANI tool [21], with the ANI value of 97.8167. Gene and functional annotations were performed with PGAP [22] v4.11 and RAS-Ttk [23]; pathways analyses were performed using the KEGG database [24] Rel 93.0 and the eggNOG mapper [25] vs 2.0.0. smBGCs were identified with antiSMASH [26] v5.0. Genome map was drawn with CGView [27] and presented in data file 6 [28].

In summary, the complete genome sequence of *H. alvei* A23BA is 4,772,047 bp in size with the overall GC content of 48.77% and sequencing coverage of 256.0 x. It comprises of one circular chromosome (4,687,005 bp; GC content 48.8%) and one circular plasmid (85,042 bp; GC content 47.2%). Genomic features include 4,217 CDSs, 25 rRNA, 92 tRNA, 8 ncRNA, 30 pseudogenes and 2 CRISPRs. Thiopeptide, beta-lactone (both showing little or no homology to known smBGCs) and siderophore smBGCs were identified (data file 7) [29]. Thiopeptides and beta-lactones are known for their antibiotic and/or anticancer activities [30, 31], while siderophores are used clinically as "Trojan horse" to deliver antibiotics to antibiotic resistant bacteria [32]. Gene clusters commonly associated with bioremediation, biocontrol, environmental adaptation, and plant growth promotion were also identified (data file 8) [33]. Please see Table 1 for links to data files 1–8.

Given the quality control measures applied and results of analyses undertaken, we believe *Hafnia alvei strain A23BA chromosome, complete genome* [34] represents a high-quality dataset that would expedite the exploration of the biosynthetic and metabolic potentials of *H. alvei* A23BA and would also enrich the comparative genomics study of *H. alvei* strains.

Limitations

This dataset was generated from a hybrid assembly to ensure accuracy and completeness. Furthermore, the hybrid assembler (Unicycler) autocorrects read errors and polishes final assemblies several times to ensure accuracy. Annotations and metabolic pathway analyses were carried out with robust and validated bioinformatics tools, and smBGCs were identified with the most comprehensive genome mining tool to date. Therefore, the authors are currently unaware of any limitations of the data.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Composition of ultra-minimal substrate growth medium	Portable Document Format file (.pdf)	https://doi.org/10.6084/m9.figshare.12781193.v [9]
Data file 2	Quality distribution of Illumina reads	Joint Photographic Experts Group file (.jpg)	https://doi.org/10.6084/m9.figshare.12643961.v1 [13]
Data file 3	Basic quality statistics of GridION sequencing data	Portable Document Format file (.pdf)	https://doi.org/10.6084/m9.figshare.12781280.v1 [14]
Data file 4	Quast report	Portable Document Format file (.pdf)	https://doi.org/10.6084/m9.figshare.12781289.v1 [18]
Data file 5	Short BUSCO summary	Joint Photographic Experts Group file (.jpg)	https://doi.org/10.6084/m9.figshare.12643937.v1 [20]
Data file 6	Circular representation of *H. alvei* A23BA genome	Joint Photographic Experts Group file (.jpg)	https://doi.org/10.6084/m9.figshare.12643949.v1 [28]
Data file 7	Predicted smBGCs of interest in *H. alvei* A23BA genome	Portable Document Format file (.pdf)	https://doi.org/10.6084/m9.figshare.12781274.v1 [29]
Data file 8	Other gene clusters of interest in *H. alvei* A23BA genome	Portable Document Format file (.pdf)	https://doi.org/10.6084/m9.figshare.13309049.v1 [33]
Data set 1	Illumina and GridION sequencing reads	Fastq file (.fastq.gz)	https://identifiers.org/ncbi/insdc.sra:SRP251948 [35]
Data set 2	Genome assembly of *H. alvei* A23BA	Fasta file (.fna)	https://identifiers.org/ncbi/insdc.gca:GCF_011617105.1 [36]

Abbreviations
GC: Guanine-Cytosine; CDSs: Coding sequences; RNA: Ribonucleic acid; rRNA: Ribosomal ribonucleic acid; tRNA: Transfer ribonucleic acid; ncRNA: Non-coding ribonucleic acid; CRISPRs: Clustered regularly interspaced short palindromic repeats; DNA: Deoxyribonucleic acid; gDNA: Genomic deoxyribonucleic acid; smBGCs: Secondary metabolite biosynthetic gene clusters; ONT: Oxford Nanopore technology; HMW: High molecular weight; #N's: Average number of uncalled bases; ANI: Average nucleotide identity.

Authors' contributions
The project was conceived and designed by OKA and AJL. Data acquisition was performed by OKA. Data analysis and interpretation was performed by OKA, NHO, ADS and AJL. The project was jointly supervised by NHO, ADS and AJL. AJL was the principal investigator. The manuscript was written by OKA and revised by NHO, ADS and AJL. All authors read and approved the final manuscript.

Availability of data and materials
Data files 1–8 described in this Data note can be freely and openly accessed on Figshare (https://figshare.com/) [9, 13, 14, 18, 20, 28, 29, 33]. Data sets 1 and 2 can be freely and openly accessed on the NCBI database. Illumina and GridION reads generated have been deposited in the Sequence Read Archive under accession number SRP251948 (Data set 1) [35]. The genome assembly of *H. alvei* A23BA has been deposited in GenBank under accession number GCF_011617105.1 (Dataset 2) [36]. The BioProject accession number for the entire project is PRJNA610978. See Table 1 and references for details and links to the data.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1] School of Pharmacy and Life Sciences, Robert Gordon University, Sir Ian Wood Building, Garthdee Road, Aberdeen AB10 7GJ, Scotland. [2] Graduate School, Robert Gordon University, The Ishbel Gordon Building, Garthdee Road, Aberdeen AB10 7QE, Scotland.

References
1. Bérdy J. Bioactive microbial metabolites. J Antibiot. 2005;58:1–26. https://doi.org/10.1038/ja.2005.1.
2. Bérdy J. Thoughts and facts about antibiotics: where we are now and where we are heading. J Antibiot. 2012;65:385–95. https://doi.org/10.1038/ja.2012.27.
3. Lugtenberg B, Kamilova F. Plant-growth-promoting rhizobacteria. Annu Rev Microbiol. 2009;63:541–56. https://doi.org/10.1146/annurev.micro.62.081307.162918.
4. Rodríguez LA, Vivas J, Gallardo CS, Acosta F, Barbeyto L, Real F. Identification of *Hafnia alvei* with the MicroScan WalkAway system. J Clin Microbiol. 1999;37:4186–8. https://doi.org/10.1128/JCM.37.12.4186-4188.1999.
5. Legrand R, Lucas N, Dominique M, Azhar S, Deroissart C, Le Solliec MA, et al. Commensal *Hafnia alvei* strain reduces food intake and fat mass in obese mice-a new potential probiotic for appetite and body weight management. Int J Obes. 2020;44:1041–51. https://doi.org/10.1038/s41366-019-0515-9.
6. Günthard H, Pennekamp A. Clinical significance of Extraintestinal *Hafnia alvei* isolates from 61 patients and review of the literature. Clin Infect Dis. 1996;22:1040–5. https://doi.org/10.1093/clinids/22.6.1040.

Awolope *et al. BMC Res Notes*   (2021) 14:8

Page 4 of 4

7. Tian B, Moran NA. Genome sequence of *Hafnia alvei* bta3_1, a bacterium with antimicrobial properties isolated from honey bee gut. Genome Announc. 2016;4:e00439-16. https://doi.org/10.1128/genomeA.00439-16.

8. Yin Z, Yuan C, Du Y, Yang P, Qian C, Wei Y, et al. Comparative genomic analysis of the *Hafnia* genus reveals an explicit evolutionary relationship between the species *alvei* and *paralvei* and provides insights into pathogenicity. BMC Genomics. 2019;20:768. https://doi.org/10.1186/s12864-019-6123-1.

9. Data File 1: Composition of ultra-minimal substrate growth medium. Figshare. 2020. https://doi.org/10.6084/m9.figshare.12781193.v1.

10. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.

11. Andrews S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

12. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. https://doi.org/10.1093/bioinformatics/btw354.

13. Data file 2: Quality distribution of Illumina reads. Figshare https://doi.org/10.6084/m9.figshare.12643961.v1 (2020).

14. Data file 3: Basic quality statistics of GridION sequencing data. Figshare https://doi.org/10.6084/m9.figshare.12781280.v1 (2020).

15. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666–9. https://doi.org/10.1093/bioinformatics/bty149.

16. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:1–22. https://doi.org/10.1371/journal.pcbi.1005595.

17. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5. https://doi.org/10.1093/bioinformatics/btt086.

18. Data file 4: Quast report. Figshare https://doi.org/10.6084/m9.figshare.12781289.v1. (2020).

19. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.

20. Data file 5: Short BUSCO summary. Figshare. 2020. https://doi.org/10.6084/m9.figshare.12643937.v1.

21. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114. https://doi.org/10.1038/s41467-018-07641-9.

22. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614–24. https://doi.org/10.1093/nar/gkw569.

23. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365. https://doi.org/10.1038/srep08365.

24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.

25. Huerta-Cepas J, Forslund K, Coelh LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34:2115–22. https://doi.org/10.1093/molbev/msx148.

26. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 50: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47:W81–7. https://doi.org/10.1093/nar/gkz310.

27. Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using the CGView family of tools. Brief Bioinform. 2019;20:1576–82. https://doi.org/10.1093/bib/bbx081.

28. Data file 6: Circular representation of *H. alvei* A23BA genome. Figshare. https://doi.org/10.6084/m9.figshare.12643949.v1 (2020).

29. Data file 7: Predicted smBGCs of interest in *H. alvei* A23BA genome. Figshare. https://doi.org/10.6084/m9.figshare.12781274.v1 (2020).

30. Just-Baringo X, Albericio F, Álvarez M. Thiopeptide antibiotics: retrospective and recent advances. Mar Drugs. 2014;12:317–51. https://doi.org/10.3390/md12010317.

31. Robinson SL, Christenson JK, Wackett LP. Biosynthesis and chemical diversity of β-lactone natural products. Nat Prod Rep. 2019;36:458–75. https://doi.org/10.1039/c8np00052b.

32. Saha M, Sarkar S, Sarkar B, Sharma BK, Bhattacharjee S, Tribedi P. Microbial siderophores and their potential applications: a review. Environ Sci Pollut Res Int. 2016;23:3984–99. https://doi.org/10.1007/s11356-015-4294-0.

33. Data file 8: Other gene clusters of interest in *H. alvei* A23BA genome. Figshare. https://doi.org/10.6084/m9.figshare.12781181.v1 (2020).

34. Awolope, OK, Di Salvo A, O'Driscoll NH, Lamb AJ. *Hafnia alvei* strain A23BA chromosome, complete genome. GenBank. https://identifiers.org/insdc:CP050150 (2020).

35. National Center for Biotechnology Information. Sequence Read Archive. https://identifiers.org/ncbi/insdc.sra:SRP251948 (2020).

36. National Center for Biotechnology Information. Assembly. https://identifiers.org/ncbi/insdc.gca:GCF_011617105.1 (2020).

Publisher's Note