nature research

Corresponding author(s): Paul G. Becher and Klas Flärdh

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Statistical analyses: Analyses were carried out in R (v. 3.3.3; R Foundation for Statistical Computing, Vienna, AT). GC-MS analysis: Before verification with reference compounds, bacterial volatiles were tentatively identified by matching their mass spectra with those in the MS Libraries (NIST 11, Wiley) using the software ChemStation (MSD Chemstation D.01.02.16 Agilent Technologies). Electroantennography (EAG and EAD): Antennal signals were recorded using a GC-EAD software (GC-EAD 2000, version 1.2.3, Syntech). Transcription Profiling: Affymetrix Arrays were scanned as described by the manufacturer and further processed by software on the scanner. This resulted in .cel files. ChIP-Seq: ChIP and total genomic DNA was sequenced by a sequencing contractor and reads received as fastq files.
Data analysis	All analysis was carried out on the Linux (Fedora 27) platform. Bash and Perl scripts were used for general purpose file manipulations. Statistical data analysis was carried out on the most recent version of R and BioConductor available at the time.

Transcription Profiling:
Data in .cel files was normalized using RMA as implemented in the Bioconductor package "affy" of R. Replicates were combined using the lmFit and eBayes functions of the Bioconductor package "limma". The resulting expression value (log2) were written to a text file.

ChIP-Seq:
R BioConductor packages Rsubread and edgeR were used for ChIP-Seq data analysis.

GC-MS analyses:
Before verification with reference compounds, bacterial volatiles were tentatively identified by matching their mass spectra with those in the MS Libraries (NIST 11, Wiley) using the software ChemStation (MSD Chemstation D.01.02.16 Agilent Technologies).

Statistics:
Statistical procedures were conducted using R version 3.3.3 and packages specified in the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The transcriptional profiling data from Affymetrix arrays and the ChIP-Seq data in Fig. 3 have been deposited at the ArrayExpress Archive of Functional Genomics Data (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5853 for the whiH mutant array data, E-MTAB-2716 for the bldM mutant array data, E-MTAB-6702 for the WhiH ChIP-Seq data, and E-MTAB-2698 for the BldM ChIP-Seq data). Source data for Figs. 1a,b,e,f,g and 4b,c and Extended Data Figs. 1a,b,c, 4, 5 and 6a,b,c are included in this article and its Supplementary Information files. Other data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Pre-study estimations were done for all tests based on simulation of hypothetical data sets, based on expected data variability and based on experience/previous studies. For analysis of data, test assumptions like data independence, distribution and variances were considered.
Data exclusions	None. Few samples in field and glass house studies could not be analysed (no data) as they got damaged by irrigation or lost for other reasons. One sample also got lost in the laboratory dispersal assay (springtails escaped). Lost samples are mentioned in source data files.
Replication	Biological triplicates for Affymetrix. Biological duplicates for ChIP-Seq. Replication is specifically stated for all other experiments.
Randomization	For field and glasshouse tests, a random number generator was used to decide upon the location of the treatments in the plots; plot and date treated as random factors in the analyses. For all laboratory experiments with springtails, animals were selected randomly from the culture. For EAG testing, test compound stimuli were randomized.
Blinding	CFU counts for different strains applied in springtail dispersal assays were done blind with respect to strain phenotype (with/without voaltile release). For the Y-tube assays data collection could not be blinded because of the nature of the assay. For the large-scale microarray adn ChIP-seq analyses, blinding is irrelevant for data analysis.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Research animals
☒	☐ Human research participants

Antibodies

Antibodies used

For anti-FLAG-WhiH ChIP-seq, protein-DNA complexes were immunoprecipitated using ANTI-FLAG® M2 Affinity Gel (Sigma cat # A220). For BldM ChIP-seq, protein-DNA complexes were immunoprecipitated with a custom made anti-BldM polyclonal antibody, raised for us by Cambridge Biochemicals Ltd using purified BldM protein provided by us.

Validation

The use of ANTI-FLAG® M2 Affinity Gel (Sigma cat # A220) in ChIP-Seq experiments with S. venezuelae is well established and has been validated in a series of previous publications (Bush MJ, et al., 2013. mBio 4:e00684-00613; Bush MJ, et al., 2016. mBio 7:e00523-00516; Bush MJ, et al., 2017. Mol. Microbiol. 104:700-711.
Similarly, the use of specific batch of anti-BldM polyclonal antibody has been established and validated in Al-Bassam MM, et al., 2014. PLoS Genet. 10:e1004554.

Research animals

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Animals/animal-derived materials

Only springtails and insects were tested or sampled in this study. Insects are legally not considered as test animals. For laboratory experiments we used females (parthenogenetic species) of the springtail Folsomia candida. Trap catches of insects, springtails and spiders from the field and greenhouse experiments were not determined to species level.

Method-specific reporting

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

WhiH:
Affymetrix data: http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5853
ChIP-Seq data: http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6702

BldM:
Affymetrix data: http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2716/
ChIP-Seq data: http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2698

Files in database submission

E-MTAB-5853.processed.1.zip
E-MTAB-5853.raw.1.zip
E-MTAB-5853.idf.txt
E-MTAB-5853.sdrf.txt
E-MTAB-5853.idf.txt_original

E-MTAB-6702.processed.1.zip
E-MTAB-6702.processed.2.zip
E-MTAB-6702.idf.txt
E-MTAB-6702.sdrf.txt
E-MTAB-6702.idf.txt_original

E-MTAB-2716.processed.1.zip
E-MTAB-2716.raw.1.zip
E-MTAB-2716.idf.txt

E-MTAB-2716.sdrf.txt

E-MTAB-2698.processed.1.zip
E-MTAB-2698.idf.txt
E-MTAB-2698.sdrf.txt
For the reads of ChIP-Seq data links are provided to the ENA (European Nucleotide Archive) on the ArrayExpress web pages for these submissions.

Genome browser session
(e.g. UCSC)

Genomic sequence fasta file, genome features bed file and ChIP peak locations bedgraph file have been provided as supplementary material. These can be used to browse the locations of the ChIP peaks on a genome browser such as IGB or IGV.

Methodology

Replicates

Two biological replicates of immunoprecipitated and non-immunoprecipitated samples.

Sequencing depth

Approximately 25 million single-ended 100 nucleotides long reads for each sample. Overall alignment rate using bowtie2 was above 95 percent in all cases. High quality sequencing depth > 250 in all cases.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

bowtie2 version 2.2.9 at default settings was used for read mapping to a bowtie2 index made from the genome nucleotide sequence of S. venezuelae. The genome sequence was divided into 25 nucleotide segments and the number of reads mapping to each of these segments were counted using the featureCounts() function of the BioConductor package Rsubread. Differential analysis of the read counts was carried out using functions calcNormFactors(), estimateDisp() and exactTest() (in this order) provided in the BioConductor package edgeR.

Data quality

Reads with mapping quality of less than 40 were excluded from analysis. There are 24 peaks above 5-fold enrichment at FDR 5%.

Software

100 nucleotide single ended reads were generated on an Illumina HiSeq 2000 sequencer. Bowtie2 version 2.2.9 was used to align reads to the S. venezuelae genomic nucleotide sequence. BioConductor packages Rsubread version 1.30 and edgeR version 3.2.2 were used for counting and differential analysis of read counts mapping to 25 nucleotide sections of the genome. Gnu Parallel (https://www.gnu.org/software/parallel) was used to run commands in parallel when possible.