Methods and Results

(in Computational Biology and Bioinformatics)

Dr Leighton Pritchard

leighton.pritchard@strath.ac.uk

https://sipbs-compbiol.github.io/

@SCompBiol

Learning objectives

- To understand the purpose of the Methods section in a thesis, report, or paper
- To understand the purpose of the Results section in a thesis, report, or paper
- To be able to recognize and identify examples of good practice in Methods and Results sections
- To be able to report bioinformatics and computational biology methods
- To be able to report bioinformatics and computational biology results

Methods sections

- You have seen Methods sections in papers (critical analysis, projects, etc.)
  - What have you seen that makes a Methods section good or useful?
  - What have you seen that makes a Methods section less useful?
  - Who reads it, and why?

Methods sections

- You have seen Methods sections in papers (critical analysis, projects, etc.)
  - What have you seen that makes a Methods section good or useful?
  - What have you seen that makes a Methods section less useful?
  - Who reads it, and why?

- **GOOD: Reproducible; specificity – materials, tools, reagents, etc.; logical order**
- **NOT SO GOOD: Referencing other papers for methods; not explaining technical terms/acronyms**
- **WHO READS?: (not everyone); people looking to reproduce the work**

What is a Methods section for?

- **It is the information that allows the reader to judge the validity of the work.**
- **Explains the procedures used to obtain the results that are presented**
  - Clearly
  - Concisely
  - Reproducibly (for a scientist competent in the area)
  - The minimal amount of information for a scientist to be able to obtain your result (± acceptable local differences)
- **The Methods section should cover all your presented Results**
  - (but don’t describe work that you’re not reporting)

Castro et al. (2022)

We downloaded all the *A. baumannii* complete genomes available at the NCBI FTP site through August 26, 2020, files containing chromosome sequences were discarded and those with plasmid sequences were kept. From these files we extracted the replication initiation proteins and their genes using the following keywords: "replication protein", "plasmid replication initiator", "plasmid replication", "DNA replication", "plasmid replicase", "replication a", "replication b", "replication c", "RepB", "rolling circle", "replication initiation", and "replicase", taking care to discard partitioning proteins, recombinases and DNA-binding proteins that are sometimes mistakenly annotated as replication proteins (Salgado-Camargo et al., 2020) (see Supplemental Materials 1 and 2 for both lists). Then, we used NCBI's batch conserved domain search (https://doi.org/10.1093/nar/gku122) to assign Pfam domains to the replicase proteins. Some *A. baumannii* plasmid sequences that were uploaded to NCBI without annotation were annotated using PROKKA, and we used the same protocol described above to identify Rep coding genes within these sequences (Seemann, 2014). We eliminated one plasmid from our collection (CP027180) because of obvious problems in its assembly.

- Is this clear?
- Is this concise?
- Could a competent scientist reproduce the work?

Castro et al. (2022)

We downloaded all the *A. baumannii* complete genomes available at the NCBI FTP site through August 26, 2020, files containing chromosome sequences were discarded and those with plasmid sequences were kept. From these files we extracted the replication initiation proteins and their genes using the following keywords: "replication protein", "plasmid replication initiator", "plasmid replication", "DNA replication", "plasmid replicase", "replication a", "replication b", "replication c", "RepB", "rolling circle", "replication initiation", and "replicase", taking care to discard partitioning proteins, recombinases and DNA-binding proteins that are sometimes mistakenly annotated as replication proteins (Salgado-Camargo et al., 2020) (see Supplemental Materials 1 and 2 for both lists). Then, we used NCBI's batch conserved domain search (https://doi.org/10.1093/nar/gku122) to assign Pfam domains to the replicase proteins. Some *A. baumannii* plasmid sequences that were uploaded to NCBI without annotation were annotated using PROKKA, and we used the same protocol described above to identify Rep coding genes within these sequences (Seemann, 2014). We eliminated one plasmid from our collection (CP027180) because of obvious problems in its assembly.

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**

Castro et al. (2022)

We downloaded all the *A. baumannii* complete genomes available at the NCBI FTP site through August 26, 2020, files containing chromosome sequences were discarded and those with plasmid sequences were kept. From these files we extracted the replication initiation proteins and their genes using the following keywords: "replication protein", "plasmid replication initiator", "plasmid replication", "DNA replication", "plasmid replicase", "replication a", "replication b", "replication c", "RepB", "rolling circle", "replication initiation", and "replicase", taking care to discard partitioning proteins, recombinases and DNA-binding proteins that are sometimes mistakenly annotated as replication proteins (Salgado-Camargo et al., 2020) (see Supplemental Materials 1 and 2 for both lists). Then, we used NCBI's batch conserved domain search (https://doi.org/10.1093/nar/gku122) to assign Pfam domains to the replicase proteins. Some *A. baumannii* plasmid sequences that were uploaded to NCBI without annotation were annotated using PROKKA, and we used the same protocol described above to identify Rep coding genes within these sequences (Seemann, 2014). We eliminated one plasmid from our collection (CP027180) because of obvious problems in its assembly.

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**
- NCBI GenBank or NCBI RefSeq?
- What if a file had chromosome and plasmid sequences?
- Which files did they download?
  - .fna? .gbff? .gff3?
- What data field did they look for keywords in?
  - Gene name? annotation?
  - What about "not a replication protein"?

Castro et al. (2022)

We downloaded all the *A. baumannii* complete genomes available at the NCBI FTP site through August 26, 2020, files containing chromosome sequences were discarded and those with plasmid sequences were kept. From these files we extracted the replication initiation proteins and their genes using the following keywords: "replication protein", "plasmid replication initiator", "plasmid replication", "DNA replication", "plasmid replicase", "replication a", "replication b", "replication c", "RepB", "rolling circle", "replication initiation", and "replicase", taking care to discard partitioning proteins, recombinases and DNA-binding proteins that are sometimes mistakenly annotated as replication proteins (Salgado-Camargo et al., 2020) (see Supplemental Materials 1 and 2 for both lists). Then, we used NCBI's batch conserved domain search (https://doi.org/10.1093/nar/gku122) to assign Pfam domains to the replicase proteins. Some *A. baumannii* plasmid sequences that were uploaded to NCBI without annotation were annotated using PROKKA, and we used the same protocol described above to identify Rep coding genes within these sequences (Seemann, 2014). We eliminated one plasmid from our collection (CP027180) because of obvious problems in its assembly.

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**

Castro et al. (2022)

We downloaded all the *A. baumannii* complete genomes available at the NCBI FTP site through August 26, 2020, files containing chromosome sequences were discarded and those with plasmid sequences were kept. From these files we extracted the replication initiation proteins and their genes using the following keywords: "replication protein", "plasmid replication initiator", "plasmid replication", "DNA replication", "plasmid replicase", "replication a", "replication b", "replication c", "RepB", "rolling circle", "replication initiation", and "replicase", taking care to discard partitioning proteins, recombinases and DNA-binding proteins that are sometimes mistakenly annotated as replication proteins (Salgado-Camargo et al., 2020) (see Supplemental Materials 1 and 2 for both lists). Then, we used NCBI's batch conserved domain search (https://doi.org/10.1093/nar/gku122) to assign Pfam domains to the replicase proteins. Some *A. baumannii* plasmid sequences that were uploaded to NCBI without annotation were annotated using PROKKA, and we used the same protocol described above to identify Rep coding genes within these sequences (Seemann, 2014). We eliminated one plasmid from our collection (CP027180) because of obvious problems in its assembly.

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**
- *Which A. baumanii* sequences without annotation?
  - all of them?
- PROKKA?
  - Where was PROKKA obtained from?
  - Which version of PROKKA?
  - Which options were used with PROKKA?

What would I want to see instead?

- Downloaded all available *A. baumanii* genomes **from NCBI RefSeq** on [date] **as .gbff files**
- **Only replicons annotated as plasmids were retained**
- **All *A. baumanii* plasmid sequences having no .gbff file feature annotation** were annotated with PROKKA **[version number and citation] using default parameters to produce corresponding annotation .gbff files**
- Extracted **from each plasmid .gbff file** all genes annotated with [list of keywords] **in their "product" field**
- **[Link to downloadable/runnable workflow]**

Hugouvieux-Cotte-Pattat et al. (2021)

A multigene maximum-likelihood phylogenetic reconstruction was performed on same set of genomes (Fig. 3). To ensure consistency of annotation between genomes, all sequences were reannotated using prodigal version 2.6.3 [43] to obtain a predicted proteome. In total 1201 single-copy orthologues were identified across the predicted proteomes of all 49 genomes, using orthofinder version 2.5.2 [44]. The protein sequences for these genes were aligned using MAFFT version 7.480 [45] and the corresponding nucleotide coding sequences threaded using t-coffee version 12.00.7fb08c2 [46]. The threaded DNA sequences were concatenated to generate one sequence per genome using the Python script concatenate_cds.py, which also generated a partition file (one partition per gene).

- Is this clear?
- Is this concise?
- Could a competent scientist reproduce the work?

Four supplementary tables and four supplementary figures are available with the online version of this article. Scripts and data enabling reproduction of the phylogenomic analysis presented in this manuscript can be obtained at https://widdowquinn.github.io/SI_Hugouvieux-Cotte-Pattat_2021/.

Hugouvieux-Cotte-Pattat et al. (2021)

A multigene maximum-likelihood phylogenetic reconstruction was performed on same set of genomes (Fig. 3).

To ensure consistency of annotation between genomes, all sequences were reannotated using prodigal version 2.6.3 [43] to obtain a predicted proteome. In total 1201 single-copy orthologues were identified across the predicted proteomes of all 49 genomes, using orthofinder version 2.5.2 [44]. The protein sequences for these genes were aligned using MAFFT version 7.480 [45] and the corresponding nucleotide coding sequences threaded using t-coffee version 12.00.7fb08c2 [46]. The threaded DNA sequences were concatenated to generate one sequence per genome using the Python script concatenate_cds.py, which also generated a partition file (one partition per gene).

Four supplementary tables and four supplementary figures are available with the online version of this article. Scripts and data enabling reproduction of the phylogenomic analysis presented in this manuscript can be obtained at https://widdowquinn.github.io/SI_Hugouvieux-Cotte-Pattat_2021/.

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**

Hugouvieux-Cotte-Pattat et al. (2021)

A multigene maximum-likelihood phylogenetic reconstruction was performed on same set of genomes (Fig. 3).

To ensure consistency of annotation between genomes, all sequences were reannotated using prodigal version 2.6.3 [43] to obtain a predicted proteome. In total 1201 single-copy orthologues were identified across the predicted proteomes of all 49 genomes, using orthofinder version 2.5.2 [44]. The protein sequences for these genes were aligned using MAFFT version 7.480 [45] and the corresponding nucleotide coding sequences threaded using t-coffee version 12.00.7fb08c2 [46]. The threaded DNA sequences were concatenated to generate one sequence per genome using the Python script concatenate_cds.py, which also generated a partition file (one partition per gene).

Four supplementary tables and four supplementary figures are available with the online version of this article. Scripts and data enabling reproduction of the phylogenomic analysis presented in this manuscript can be obtained at https://widdowquinn.github.io/SI_Hugouvieux-Cotte-Pattat_2021/.

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**
- What software parameters were used?
- Where can I get concatenate_cds.py?

Hugouvieux-Cotte-Pattat et al. (2021)

A multigene maximum-likelihood phylogenetic reconstruction was performed on same set of genomes (Fig. 3). To ensure consistency of annotation between genomes, all sequences were reannotated using prodigal version 2.6.3 [43] to obtain a predicted proteome. In total 1201 single-copy orthologues were identified across the predicted proteomes of all 49 genomes, using orthofinder version 2.5.2 [44]. The protein sequences for these genes were aligned using MAFFT version 7.480 [45] and the corresponding nucleotide coding sequences threaded using t-coffee version 12.00.7fb08c2 [46]. The threaded DNA sequences were concatenated to generate one sequence per genome using the Python script concatenate_cds.py, which also generated a partition file (one partition per gene).

- **What are they telling us?**
- **What are they leaving out?**
  - **Does it matter?**
- **Could a competent scientist reproduce this?**
- What software parameters were used?
- Where can I get concatenate_cds.py?

Four supplementary tables and four supplementary figures are available with the online version of this article. Scripts and data enabling reproduction of the phylogenomic analysis presented in this manuscript can be obtained at https://widdowquinn.github.io/SI_Hugouvieux-Cotte-Pattat_2021/.

Bioinformatics' secret weapon

SI_Hugouvieux-Cotte-Pattat_2021

Supplementary information for pyani analyses reported in Hugouvieux-Cotte-Pattat et al. (2021) IJSEM, describing the novel genus Paradisiaca

View the Project on GitHub
widdowquinn/SI_Hugouvieux-Cotte-Pattat_2021

Reproducing analyses (quickstart)

You can use this archive to browse, validate, reproduce, or build on the phylogenomics analysis for the Hugovieux-Cotte-Pattat *et al.* (2021) manuscript. We recommend creating a conda environment specific for this activity, for example using the commands:

```
conda create --name musicola python=3.8 -y
conda activate musicola
conda install --file requirements.txt -y
```

All scripts used to generate the phylogenomic analysis are found in the `scripts/` subdirectory, and can be run in order to regenerate the analysis:

```
scripts/download_genomes.sh
scripts/annotate_genomes.sh
scripts/run_anim.sh
scripts/find_orthologues.sh
scripts/align_scos.sh
python scripts/extract_cds.py
scripts/backtranslate.sh
python scripts/concatenate_cds.py
scripts/build_tree.sh
```

Bioinformatics' secret weapon

- **We can share the exact code used to analyse our data and produce figures using services like GitHub, BitBucket, Zenodo, GitLab, FigShare, etc.**
- **We can share the data and results, too!**
- Make your code/scripts/workflows available
- Put the link to the code/script/workflow in the Methods section
- **You can then outline the methodology and refer the reader to the online resource for detail**
- You still need enough information for the reader to understand that what was done was valid, in principle

Can a figure help?

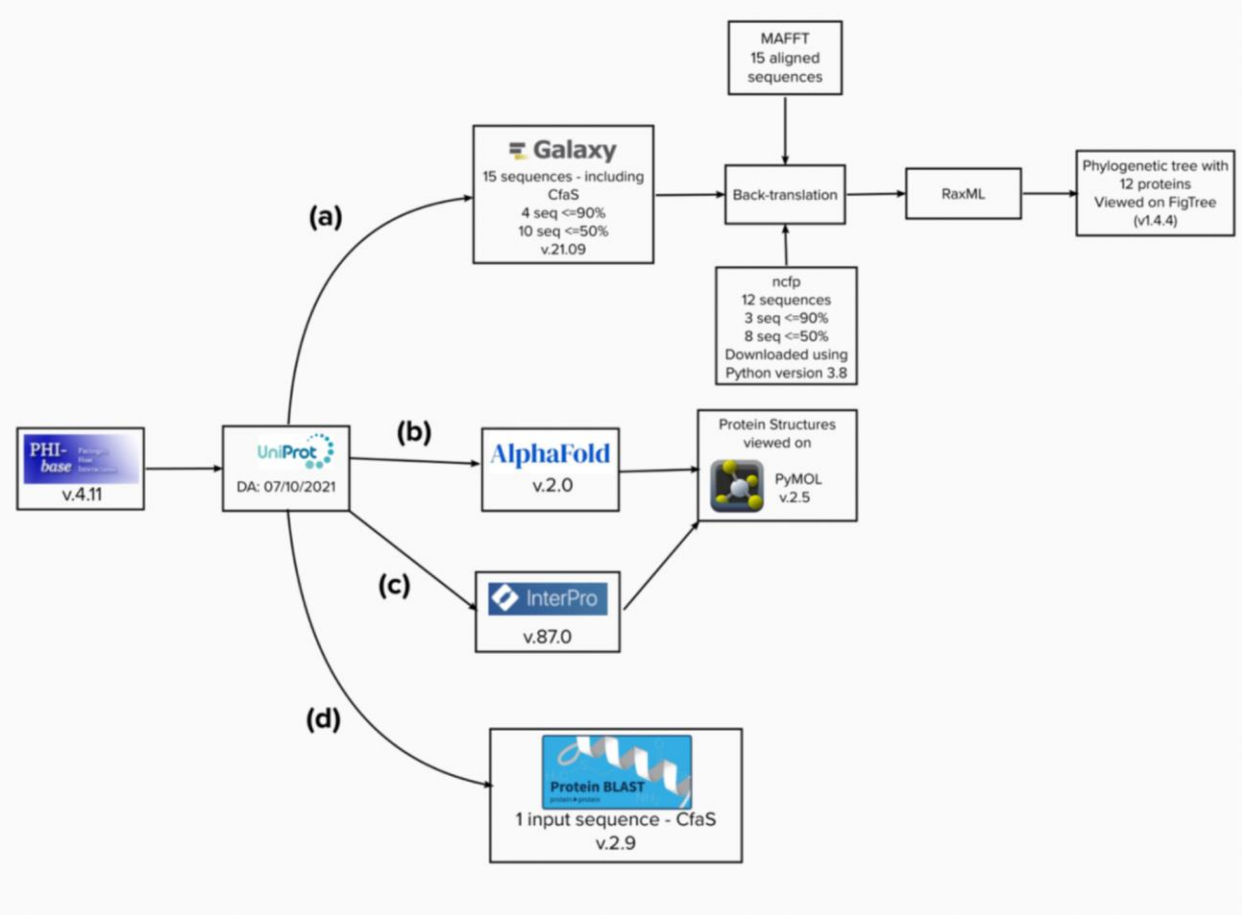
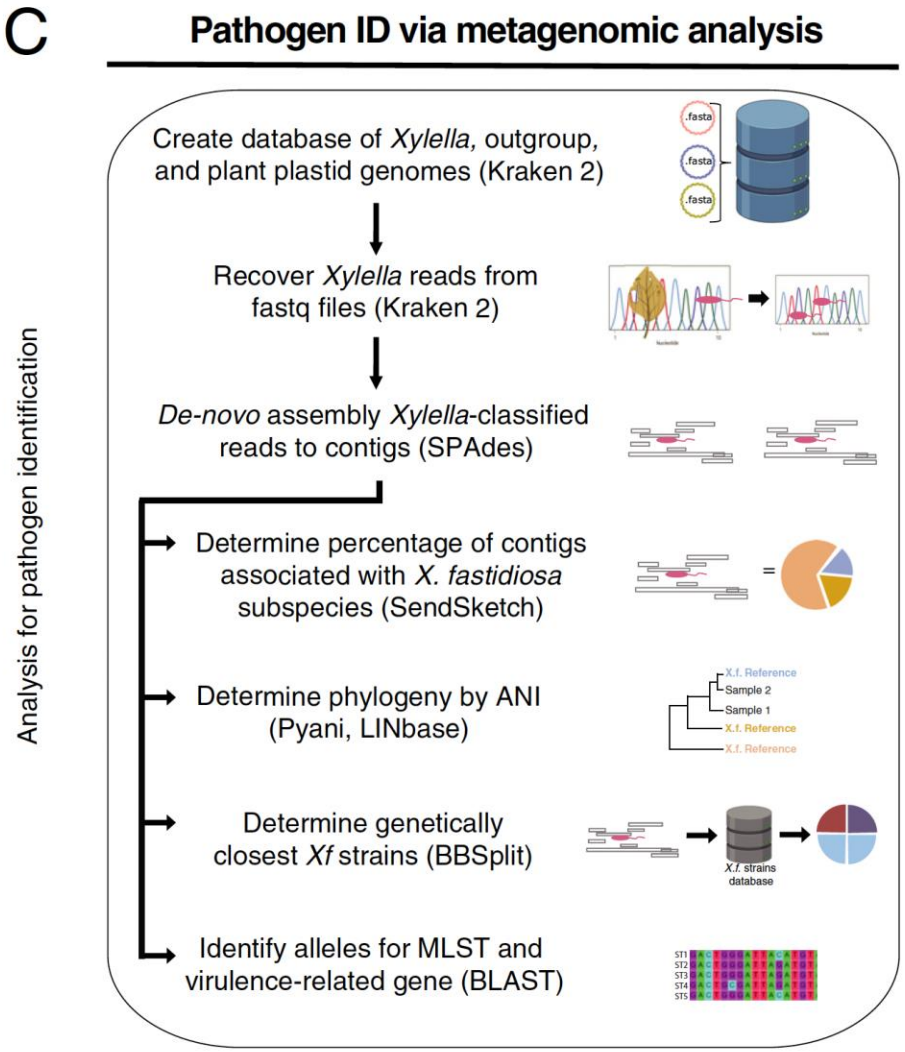


Figure 5: Methodology flow chart.
Databases accessed for (a) Producing a phylogenetic tree using tools within Galaxy and ncfp (b) Visualising 3D protein structures using PyMOL (c) Examining predicted functional domains using InterPro and (d) Investigating evolutionary history of CfaS using BLAST.

- Is this clear?
- Is this concise?
- Is this easier to produce and understand than several paragraphs of text by itself?
- **Visual communication is a "shortcut" to your reader's understanding**
- **Detail in the text, structure in the flowchart**

Can a figure help?



- Is this clear?
- Is this concise?
- Is this easier to produce and understand than several paragraphs of text by itself?
- **Visual communication is a "shortcut" to your reader's understanding**
- **Detail in the text, structure in the flowchart**

When is your Methods section finished?

- Ask yourself:
  - **"If I were reading this for the first time, and hadn't done the work myself, could I reasonably reproduce what I did from this text?"**
  - **If your answer is "no" – IT'S NOT FINISHED**
  - **If your answer is "maybe" – IT MIGHT NOT BE FINISHED**

- **If in doubt: get a friend to read it**

Results sections

- You have seen Results sections in papers (critical analysis, projects, etc.)
  - What have you seen that makes a Results section good or useful?
  - What have you seen that makes a Results section less useful?
  - Who reads it, and why?

Results sections

- You have seen Results sections in papers (critical analysis, projects, etc.)
  - What have you seen that makes a Results section good or useful?
  - What have you seen that makes a Results section less useful?
  - Who reads it, and why?

- **GOOD: state what results imply; understandable figures – (self-contained); link back to data origin; link to data availability**
- **LESS GOOD: overcomplicated figures; results that are not consequential to the research question**
- **WHO READS RESULTS?: Anyone interested in the work/research question**

What is a Results section for?

- **The Results section describes the (main) findings of your work**
  - Clearly
  - Concisely
  - In a logical order/understandable sequence
- **The Results section lays the framework for evaluation of the results in the Discussion section**
- **The Results section allows the reader to evaluate the soundness of your conclusions**

What goes into a Results section?

- **Your results, e.g.**
  - Text
  - Graphs/charts/plots
  - Tables
  - Quantitative results (e.g. statistical test output: were null hypotheses rejected?)
  - Qualitative results (e.g. trends or recurring patterns)
  - Links to online/supplementary results
- **The exact detail of what is included, and its order, depends on the scope and nature of your study, and your research question**

Olm et al. (2020)

RESULTS

Discrete sequence groups exist in all analyzed genome sets. Sets of microbial genomes without the selection biases introduced by isolation were generated from metagenomic studies of three environments: infant fecal samples (1,163 metagenomes collected from 160 hospitalized premature infants over 5 years) (33), the ocean (234 metagenomes collected from the global *Tara* Oceans Expedition over 7 years) (34), and a meadow soil ecosystem (60 metagenomes collected from three depths at five locations for five time points across a grassland meadow) (26). A taxonomically balanced set of genomes from RefSeq was generated by randomly choosing 10 genomes from each of the 480 species in RefSeq with at least 10 genomes (see Table S1 in the supplemental material; see Materials and Methods for details). All genomes within each of the four sets were compared to each other in a pairwise manner using the FastANI algorithm (10). Discrete sequence groups based on both ANI and genome alignment percentages were found in all genome sets (Fig. 1). Notably, species identity gaps were even more prominent in genome sets based on MAGs (metagenome-assembled genomes) than in those from RefSeq (which mainly consists of cultured isolate genomes). Comparisons of RefSeq genomes marked as belonging to the same bacterial species versus different bacterial species showed that the identity gap was largely consistent with annotated NCBI species taxonomy and that most genome clusters segregated from each other with a cluster boundary at around 95%. Thus, the analysis is consistent with prior suggestions that this cutoff delineates the species boundary. MAGs from the human microbiome were often very similar to each other (>98% ANI), whereas MAG clusters from the ocean included greater numbers of divergent strain types. In contrast, most of the comparisons involving genomes from soil involved distinct species.

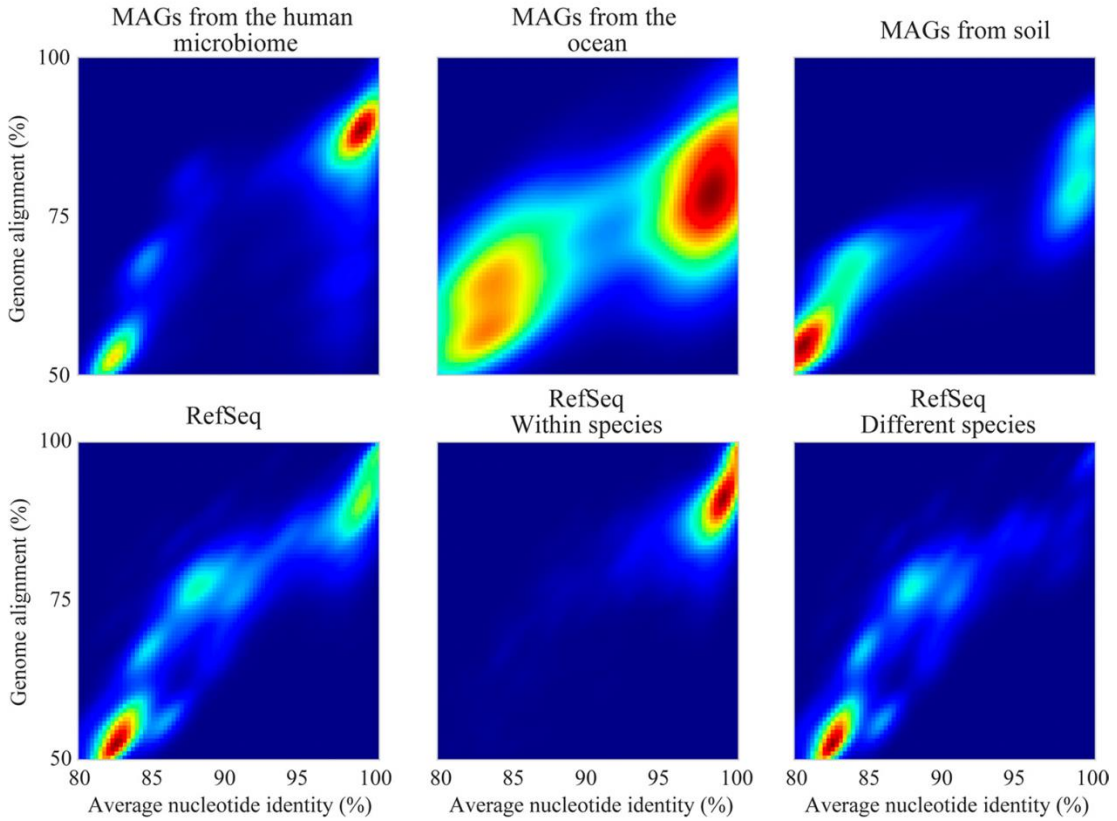


FIG 1 Average nucleotide identity gaps exist near ~95% ANI in all tested genome sets. Each plot is a histogram of average nucleotide identity and genome alignment percentage values resulting from pairwise comparison within a genome set. Higher-intensity colors represent a higher density of comparisons with that particular ANI and genome alignment percentage. The top row contains data from three sets of metagenome-assembled genomes (MAGs) from different environments. The bottom row displays data from NCBI RefSeq (rarefied to reduce taxonomic bias; see Materials and Methods), RefSeq with only comparisons between genomes annotated as the same species included, and RefSeq with only comparisons between genomes annotated as different species included.

Olm et al. (2020)

RESULTS

Discrete sequence groups exist in all analyzed genome sets. Sets of microbial genomes without the selection biases introduced by isolation were generated from metagenomic studies of three environments: infant fecal samples (1,163 metagenomes collected from 160 hospitalized premature infants over 5 years) (33), the ocean (234 metagenomes collected from the global *Tara* Oceans Expedition over 7 years) (34), and a meadow soil ecosystem (60 metagenomes collected from three depths at five locations for five time points across a grassland meadow) (26). A taxonomically balanced set of genomes from RefSeq was generated by randomly choosing 10 genomes from each of the 480 species in RefSeq with at least 10 genomes (see Table S1 in the supplemental material; see Materials and Methods for details). All genomes within each of the four sets were compared to each other in a pairwise manner using the FastANI algorithm (10). Discrete sequence groups based on both ANI and genome alignment percentages were found in all genome sets (Fig. 1). Notably, species identity gaps were even more prominent in genome sets based on MAGs (metagenome-assembled genomes) than in those from RefSeq (which mainly consists of cultured isolate genomes). Comparisons of RefSeq genomes marked as belonging to the same bacterial species versus different bacterial species showed that the identity gap was largely consistent with annotated NCBI species taxonomy and that most genome clusters segregated from each other with a cluster boundary at around 95%. Thus, the analysis is consistent with prior suggestions that this cutoff delineates the species boundary. MAGs from the human microbiome were often very similar to each other (>98% ANI), whereas MAG clusters from the ocean included greater numbers of divergent strain types. In contrast, most of the comparisons involving genomes from soil involved distinct species.

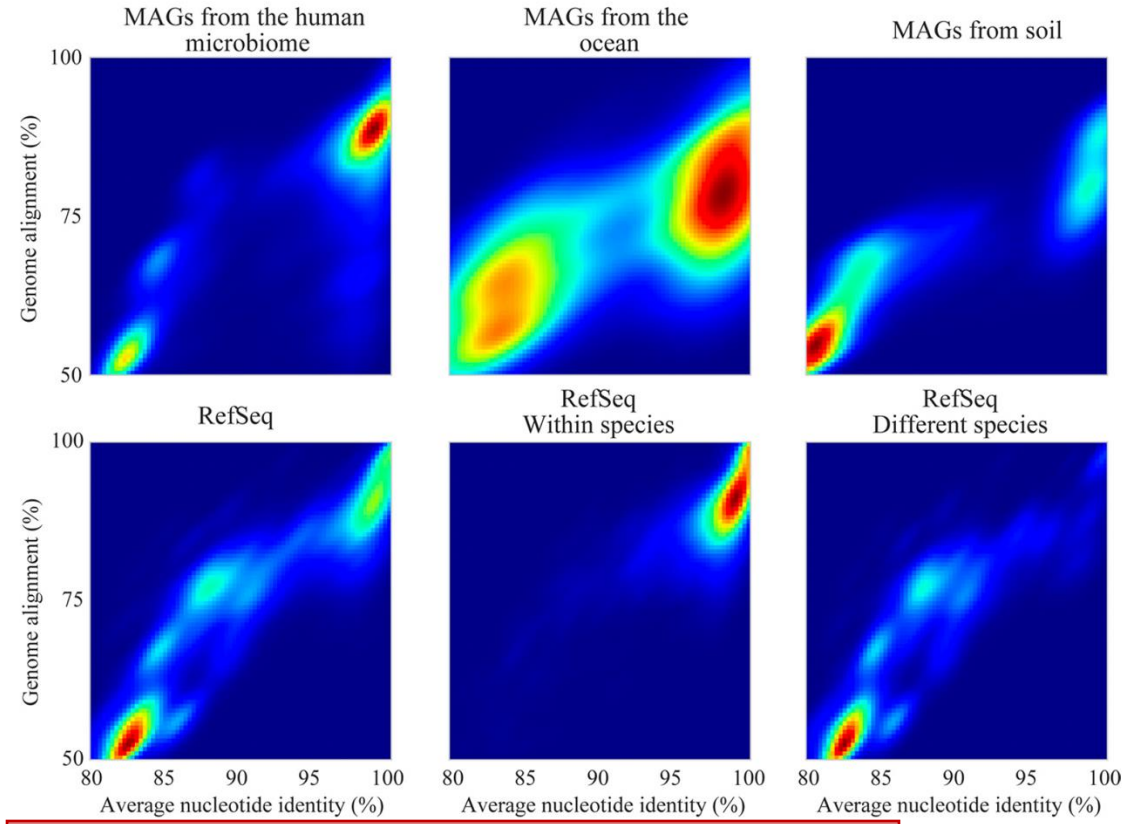


FIG 1 Average nucleotide identity gaps exist near ~95% ANI in all tested genome sets. Each plot is a histogram of average nucleotide identity and genome alignment percentage values resulting from pairwise comparison within a genome set. Higher-intensity colors represent a higher density of comparisons with that particular ANI and genome alignment percentage. The top row contains data from three sets of metagenome-assembled genomes (MAGs) from different environments. The bottom row displays data from NCBI RefSeq (rarefied to reduce taxonomic bias; see Materials and Methods), RefSeq with only comparisons between genomes annotated as the same species included, and RefSeq with only comparisons between genomes annotated as different species included.

What goes into a Results section?

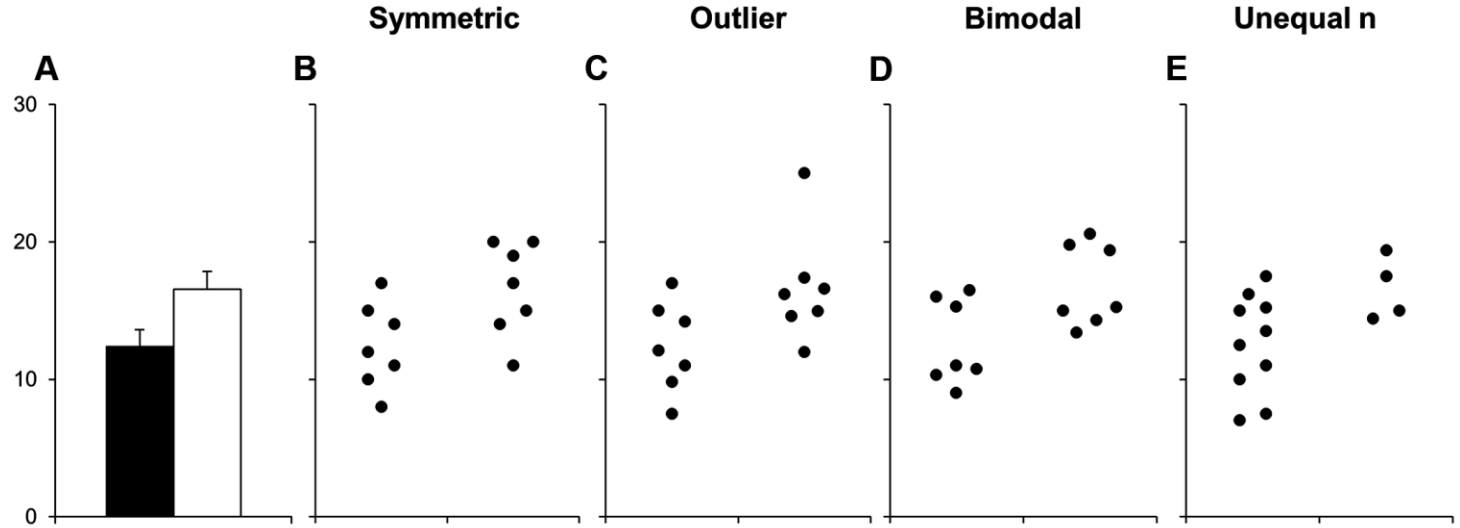
- **Restatement of the aim of the research (i.e. context)**
- **Explanation of the data obtained and findings from analyses**
  - Summaries of data (descriptive statistics), or the data itself
  - Reports of statistical or other analytical output
- **Tables, figures, and anything else that makes your results more digestible for the reader**
  - Summarise the result being shown, in your figure legend/table caption
- **Sometimes the data is clear enough you can state what it means**
- **Sometimes you need to explain what the data means, to the reader**
  - (this can be a judgement call)

What does not go in a Results section?

- **Material that is irrelevant to your research question**
  - This does not mean “negative” results – just material that has no bearing on the question you are answering/hypothesis you are testing
- **Speculation and hyperbole**
  - Research findings very rarely “prove” or “demonstrate” things: they usually only accept or reject an existing hypothesis, or suggest new hypotheses to test
  - **Null hypothesis significance tests can only accept or reject the null hypothesis. They cannot prove your alternative hypothesis!**
- **The exact detail of what is (not) included, and its order, depends on the scope and nature of your study, and your research question**

Reporting statistics (1)

- **Understand what your statistics show**
- **Are you testing a hypothesis (e.g. *t*-test)?**
  - Does your analysis accept or reject the null hypothesis?
- **Are you summarizing data (e.g. showing data points on a graph)?**
  - Show the complete data (i.e. use a 1D scatterplot instead of a bar graph)
  - Avoid bar graphs, and dynamite plots



Reporting statistics (2)

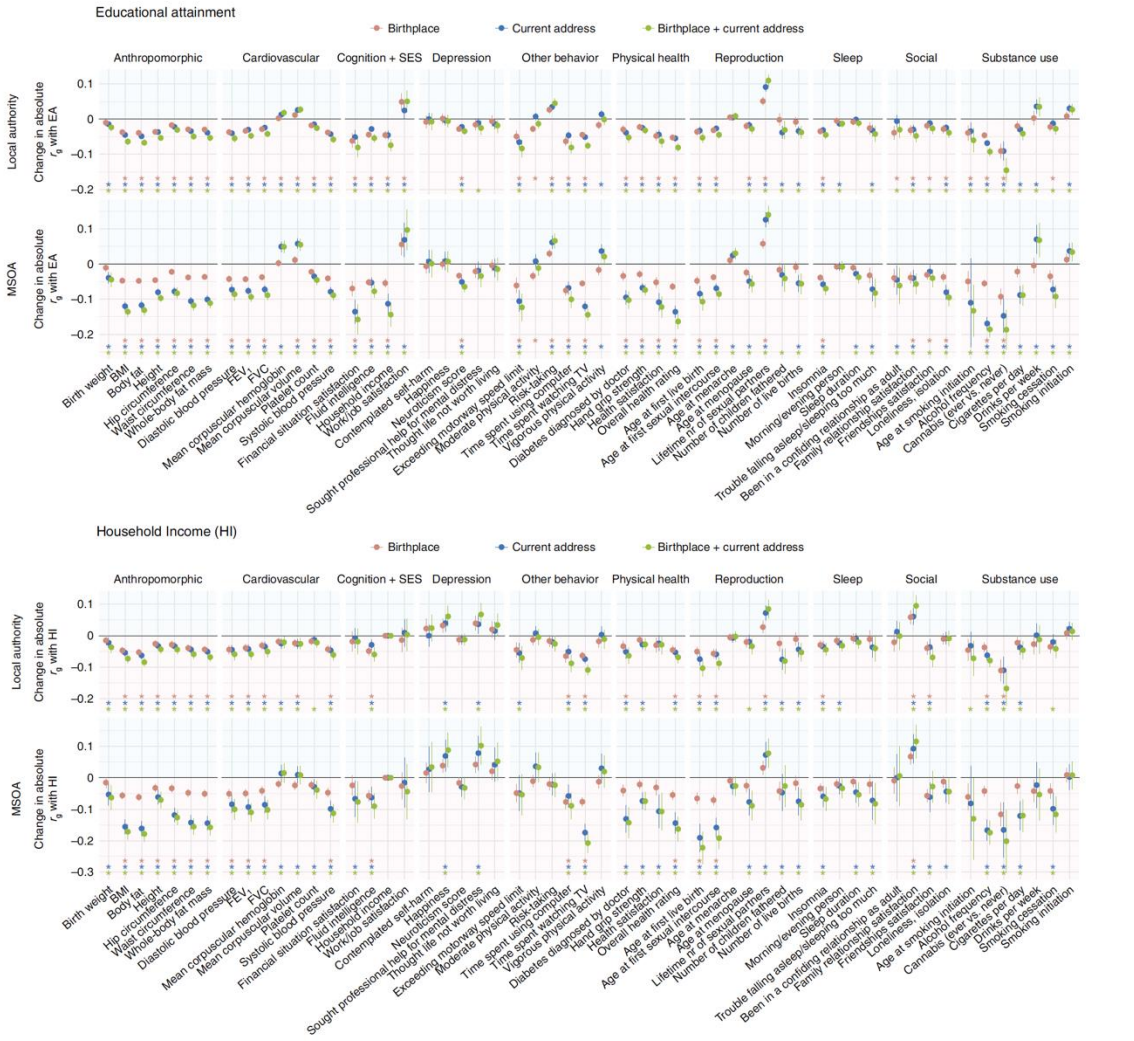
- **Report the following where relevant:**
  - Sample size (how many data points)
  - Data transformation (e.g. log transform, omission of datapoints/"outliers")
  - Statistical test: name the test and how it was implemented (e.g. R, SPSS, Excel – detail could be in Methods)
  - Directionality of the test (e.g. one-tailed or two-tailed)
  - Effect size
  - Multiple test correction (often overlooked!)
  - Your chosen level of significance (often *P*=0.05, but an arbitrary choice)
  - Confidence intervals (e.g. parameter estimation)
  - Variability of data (e.g. standard deviation, standard error of the mean)
  - Estimated *P*-value

Reporting statistics (3)

- **What you need to state depends on the test**
- **For null hypothesis significance tests in general (*t*-test, Chi-square, etc.)**
  - **DO Report *P*-values as actual values (e.g. *P*=0.03)**
  - **DO NOT report *P*-values as inequalities (e.g. *P*<0.05) unless they are sufficiently small (e.g. *P*<0.0001)**
  - **DO NOT simply report "a significant difference was found (*P*<0.05)"**
  - Your statistical significance threshold is arbitrary – **results of *P*=0.051 and *P*=0.049 should be interpreted similarly, regardless of the use of a threshold of *P*=0.05 for statistical significance**
  - *P*-values you consider non-significant should also be reported as actual values
  - Include the degrees of freedom for your tests

Abdellaoui *et al.* (2022)

Controlling for local authority or MSOA region significantly reduced the genetic correlations between most of the 56 complex traits and education and income (Fig. 6). We observed the most and strongest reductions when controlling for birthplace and current address jointly and the smaller MSOA regions. When controlling for MSOA regions based on both birthplace and current address jointly, 41 traits showed a significant reduction in the genetic correlation with educational attainment. The five most significant decreases were observed for height (from 0.20 to 0.11, $P_{change} = 9 \times 10^{-70}$), body fat percentage (from −0.34 to −0.20, $P_{change} = 1 \times 10^{-69}$), BMI (from −0.31 to −0.17, $P_{change} = 1 \times 10^{-68}$), alcohol frequency (from −0.42 to −0.23, $P_{change} = 1 \times 10^{-65}$) and time spent watching television (from −0.69 to −0.54, $P_{change} = 2 \times 10^{-63}$). When controlling for MSOA regions based on birthplace and current address, 35 traits showed a significant reduction of the genetic correlation with household income. The five most significant decreases were observed for body fat percentage (from −0.32 to −0.15, $P_{change} = 5 \times 10^{-44}$), BMI (from −0.33 to −0.15, $P_{change} = 7 \times 10^{-41}$), time spent watching television (from −0.62 to −0.41, $P_{change} = 2 \times 10^{-37}$), whole-body fat mass (from −0.25 to −0.09, $P_{change} = 3 \times 10^{-37}$) and waist circumference (from −0.29 to −0.13, $P_{change} = 5 \times 10^{-34}$), which is the same top five as for the polygenic score analyses in siblings summarized in Table 1.

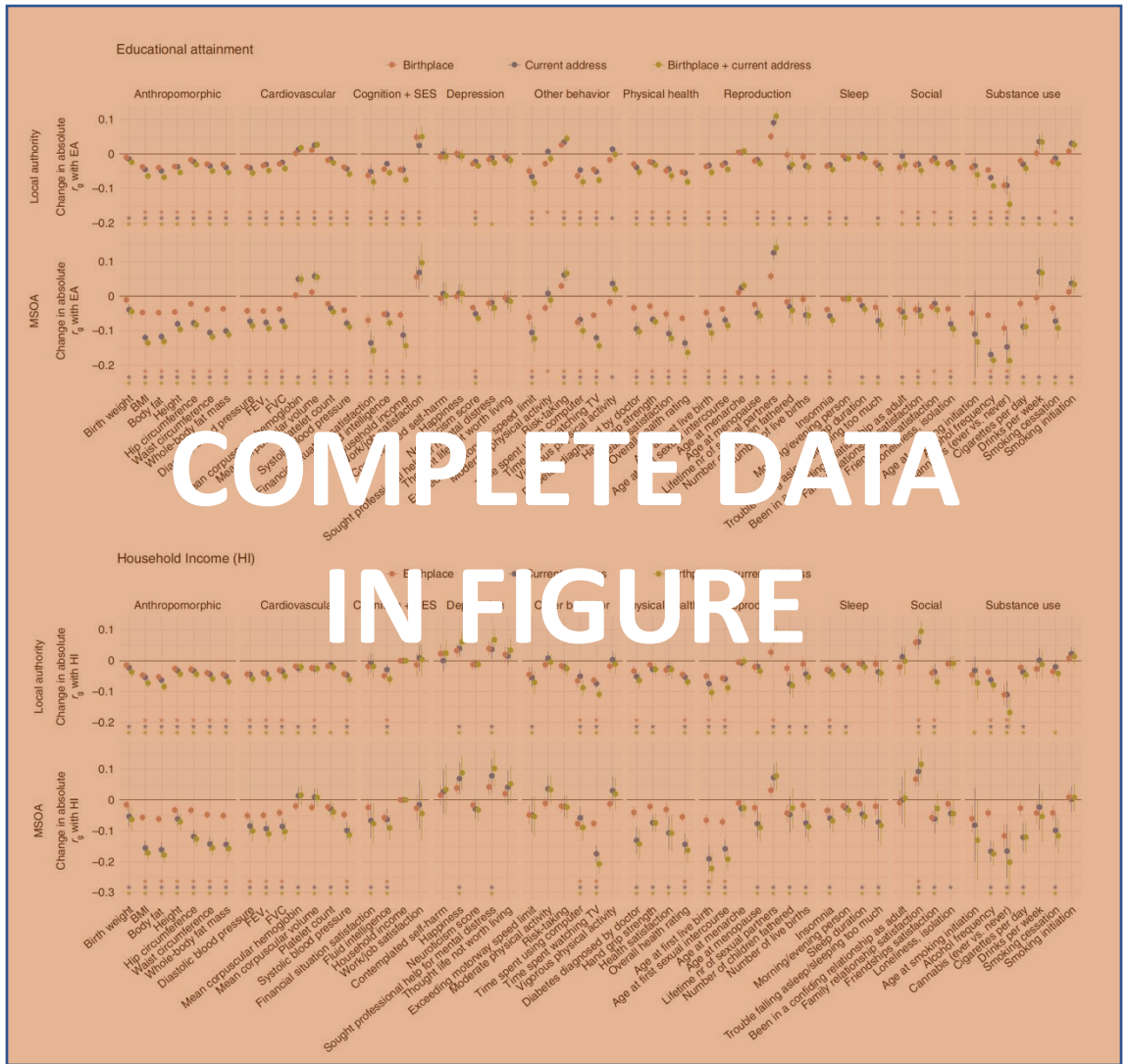


Abdellaoui *et al.* (2022)

Controlling for local authority or MSOA region significantly reduced the genetic correlations between most of the 56 complex traits and education and income (Fig. 6). We observed the most and strongest reductions when controlling for birthplace and current address jointly and the smaller MSOA regions. When controlling for MSOA regions based on both birthplace and current address jointly, 41 traits showed a significant reduction in the genetic correlation with educational attainment. The five most significant decreases were observed for height (from 0.20 to 0.11, $P_{change} = 9 \times 10^{-70}$), body fat percentage (from −0.34 to −0.18, $P_{change} = 1 \times 10^{-68}$), BMI (from −0.31 to −0.17, $P_{change} = 1 \times 10^{-68}$), alcohol frequency (from −0.42 to −0.23, $P_{change} = 1 \times 10^{-63}$) and time spent watching television (from −0.69 to −0.54, $P_{change} = 2 \times 10^{-63}$). When controlling for MSOA regions based on birthplace and current address, 35 traits showed a significant reduction of the genetic correlation with household income. The five most significant decreases were observed for body fat percentage (from −0.32 to −0.15, $P_{change} = 5 \times 10^{-44}$), BMI (from −0.33 to −0.15, $P_{change} = 7 \times 10^{-41}$), time spent watching television (from −0.62 to −0.41, $P_{change} = 2 \times 10^{-37}$), whole-body fat mass (from −0.25 to −0.09, $P_{change} = 3 \times 10^{-37}$) and waist circumference (from −0.29 to −0.13, $P_{change} = 5 \times 10^{-34}$), which is the same top five as for the polygenic score analyses in siblings summarized in Table 1.

SUMMARY

MAIN EFFECT SIZES AND P-VALUES



COMPLETE DATA IN FIGURE

In summary

- **The Methods section tells the reader how you did the work**
  - (the reader can then determine whether the study is valid)
- **The Results section tells the reader what you found**
  - (the reader can then determine whether your conclusions are sound)
- Software used should be reported with version number and citation
- All software parameters for an analysis should be made available
  - A blanket statement like “default parameters were used for all software unless noted otherwise” may be useful
- **Make your scripts/code available**

Useful links (1)

- The importance of a Methods section: https://www.enago.com/academy/importance-methods-section-academic-papers/
- How to write a Methods section: https://www.enago.com/academy/how-to-write-the-methods-section-of-a-scientific-article/
- How to write a Methods section: https://pubmed.ncbi.nlm.nih.gov/15447808/

Useful links (2)

- How to write a Results section: https://citetotal.com/writing-guides/how-to-write-a-results-section/
- How to write a Results section: https://www.scribbr.co.uk/thesis-dissertation/results-section/
- Writing up a Results section: https://www.oxbridgeessays.com/blog/writing-results-section-dissertation/

Useful links (3)

- How to report statistics: https://plos.org/resource/how-to-report-statistics/
- Guidelines for reporting statistics: https://support.jmir.org/hc/en-us/articles/360019690851-Guidelines-for-reporting-statistics
- APA statistics style: https://www.scribbr.com/apa-style/numbers-and-statistics/
- Guidelines for reporting statistics: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6397060/