

Scientific Writing

魏桐

11/9/2022

- What is scientific writing?
- Manuscript preparation
- Manuscript submission

Part I

What is scientific writing

How to get your work published

- Your research is good...
 - It has novelty.
 - The experiments is well-designed.
 - The results are solid.
 - ...
- And more importantly nowadays, your manuscript is **well organized and written!**

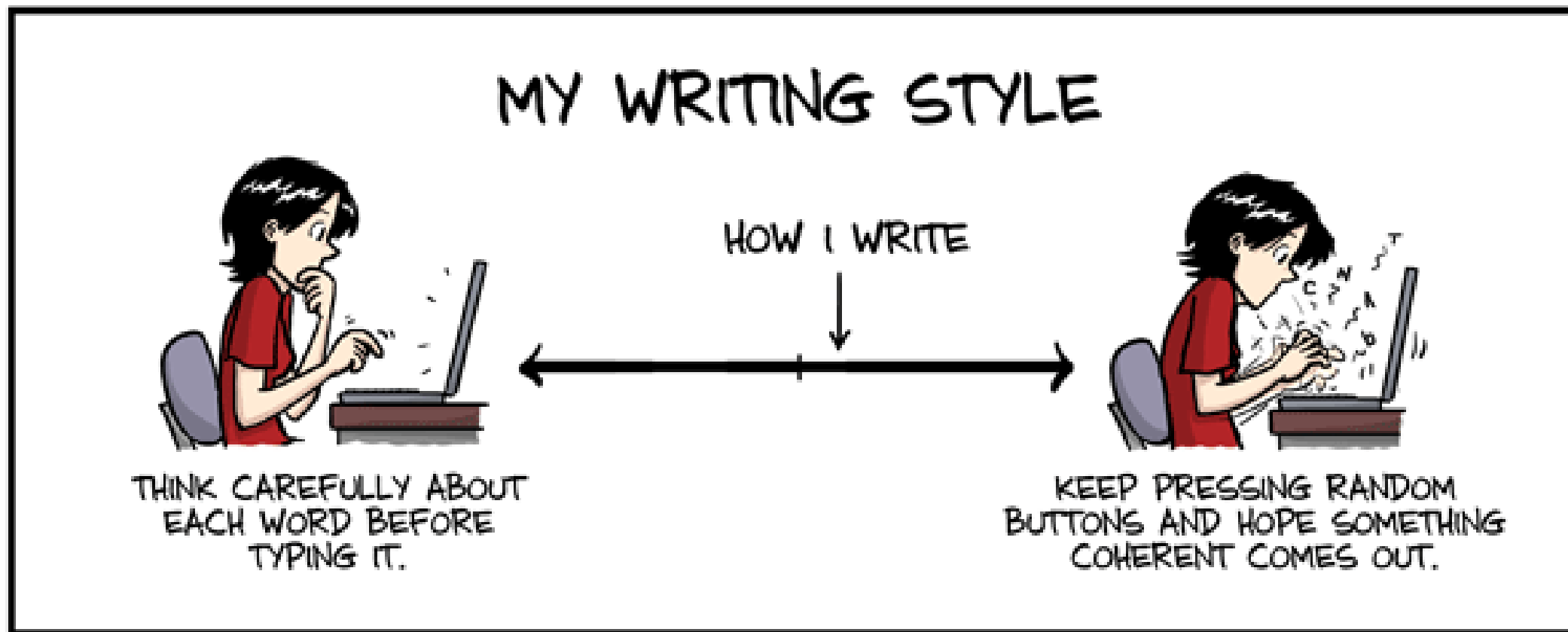
What happens to poorly written papers

- Editors **miss** the point.
- Editors feel the paper **will not pass** the reviewing process.
 - Most likely reject.
- Reviewers feel **frustrated**.
 - Likely reject.
- Reviewers feel **confused**.
 - Probably raise extra questions or ask for more results.

Writing a manuscript is
as important as doing experiments.

For high-impact journals,
writing is often more important.

Writing needs practices



JORGE CHAM © 2017

WWW.PHDCOMICS.COM

What is scientific writing?

- Scientific writing is **the technical writing** that is designed to **communicate** scientific information to other scientists.
- Depending on the specific scientific genre—a journal article, a scientific poster, or a research proposal, **the details vary but have a greater similarity**.
- Important hallmarks of all scientific writing are,
 - Has to be **logical and organized**,
 - Must be **exact and precise**,
 - Needs to be set **within the context** of published work.

Key questions related to your research

- **Why** did you do it?
- **How** did you do it?
- **What** did you get?
- **So what** did it mean?

Q&As related to your work

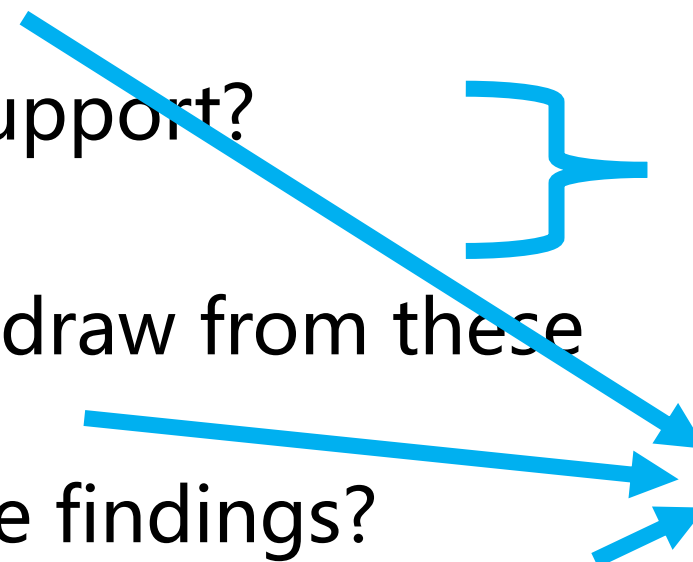
- Why is your work important?
- What problem do you want to solve?
- What is your main discovery?
- What experiments did you do to support?
- How did you do the experiments?
- What specific conclusions can you draw from these experiments?
- What can you conclude from all the findings?
- What is the significance of your discovery?

} WHY

} HOW

WHAT

SO WHAT

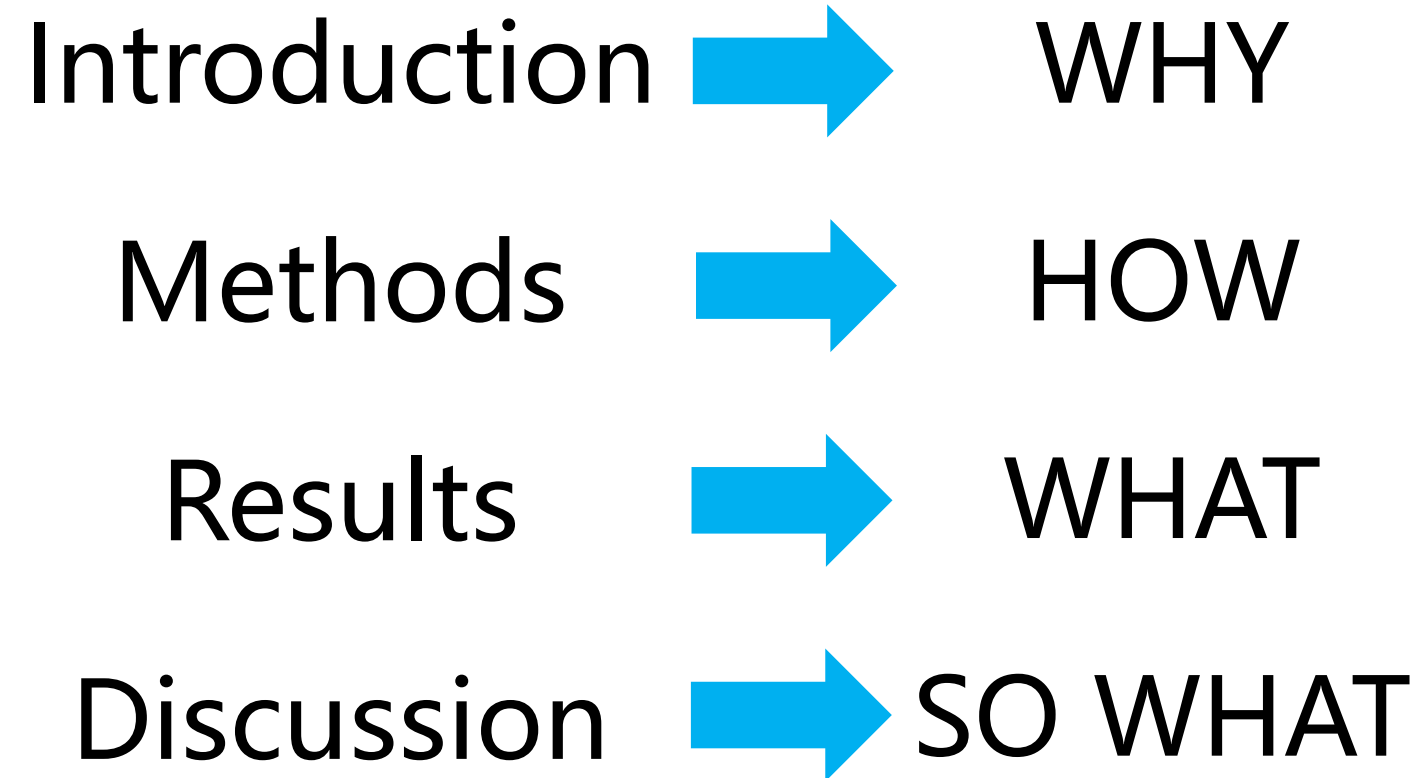


- Why did you do it?
 - ... is important in organ development/crop science; however, the mechanism remains unclear.
- How did you do it?
 - We carried out a multi-omic approach ...
- What did you get?
 - The results showed that genes were associated with ...
- So what did it mean?
 - Our work discovered key players and shed light on ...

- Why did you do it?
 - To investigate/reveal/study the mechanism of [your research], ..
 - To further explore the population structure, ...
- How did you do it?
 - We conducted a transcriptomic approach ...
 - We analyzed the SNPs from two populations ...
- What did you get?
 - The results showed that genes were differentially expressed ...
 - The population analysis revealed additional structure ..
- So what did it mean?
 - Our data indicate the transcriptomic reprogramming during ...
 - The results suggest the domestication history of ...

- Why did you do it?
 - ... is **important** to [your research field]
- How did you do it?
 - We **carried out** a omics approach ...
- What did you get?
 - Our results **revealed** that [gene/family/pathway] is involved in ..., which agrees previous studies (**citation**).
- So what did it mean?
 - Our work **demonstrate** the necessity of applying genomic approaches in...

Abstract



- **Background + question**
- **A sentence or phrases about methods**
- **Major discoveries**
- **Conclusion + significance**

Part II

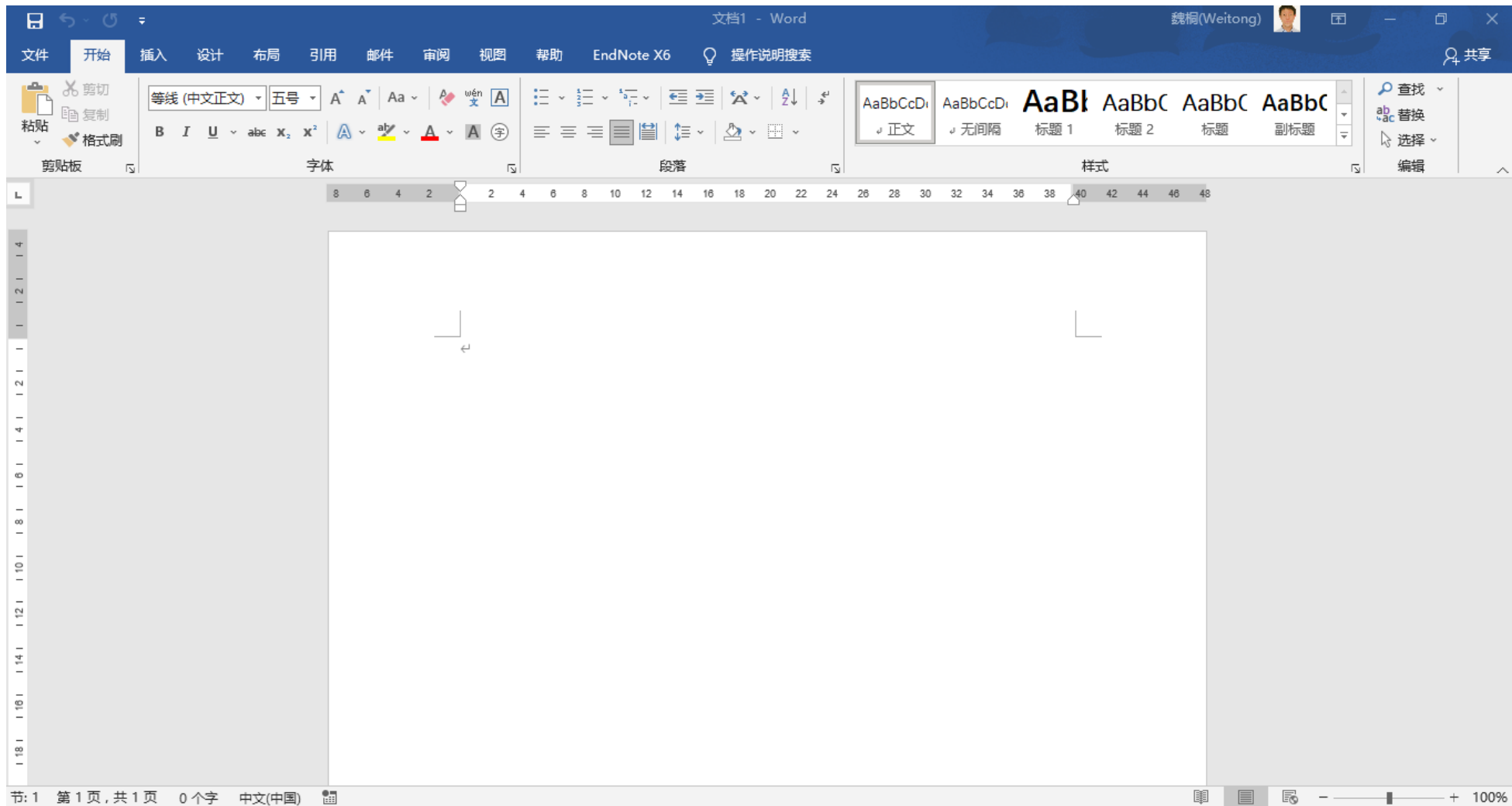
Manuscript drafting

3 phases during manuscript preparation

- The preparation phase
 - A **final set of processed data**, assembly, variants, expression, etc.
 - A complete set of **preliminary figures & tables**
- The drafting phase
 - **An story line** with logic
 - Figures + Results & Methods
 - Discussion + Introduction
 - Abstract + Title
- The editing phase
 - Seek for **professional comments**
 - Language polishing

- Get all the results together, e.g. processed results (**figures & tables if possible**)
 - Get the methods ready, i.e. software, parameters, filtering criteria, etc.
 - Scan the literature briefly
 - Revisit **the scientific question(s)**
-
- Conceptualize/visualize your work, and draw the OUTLINE

Ready to draft the manuscript!



- **OUTLINE** guides the drafting
- **BOTTOM-UP** writing
- Write out **an rough Abstract** from the outline
- Assign **a topic sentence** for each paragraph in each section
- Image **writing for people** without any background knowledge

Outcomes from the drafting phase

- **An rough abstract** with logic
- Main figure layout
- **Results & Methods**
- A rough Discussion
- A rough Introduction
- A revised Abstract if results change
- A working title

- It answers the **WHY, HOW, WHAT** and/or **SO WHAT** questions.
- It is a special tool to,
 - Guide figure preparation,
 - Dictate results writing,
 - Review the whole draft.
- **Know your data** and spent some time, say 30 minutes, to draw the outline.

WHY

1. Lettuce (*Lactuca sativa*) is one of the most important vegetables worldwide; however, the natural history and genetic basis of its domestication remain largely unknown.

HOW

2. Here in this study, we sequenced a total of 445 *Lactuca* accessions including cultivated lettuce and wild species.

WHAT

3. The phylogenetic analysis clarified the relationship between *L. sativa* and other *Lactuca* species, especially those used as breeding materials.

4. As the proposed progenitor of cultivated lettuce, *L. serriola* was categorized into five geographic groups and the one from Western Asia shared the common ancestor with the modern crop lettuce.

5. We identified genomic regions under selection during lettuce domestication and improvement, from which genes were discovered in flowering, trichome formation, stress response.

SO WHAT

6. The genome sequence and the variation map generated here not only reveals the population structure and domestication history of cultivated lettuce, but facilitates gene discovery for favorable traits in the future lettuce breeding.

- Describe all the experimental procedures **in details**
 - Experimental materials, species, ecotypes, batches, etc.
 - Treatment, chemicals, concentration, time, etc.
 - Data analysis, software versions, parameters, formats, etc.
- Write methods **after the analyses is done**.
- This is quite straightforward.

- Variant calling was carried out following the Genome Analysis ToolKit (GATK version 4.0.3.0) Best Practices [41](#). Raw reads were filtered by Trimmomatic using the same parameter as in the genome assembly procedure, and aligned to the *L. sativa* cv. Salinas reference genome (v8.0, downloaded from <http://genomevolution.org/coge>)¹⁴ using BWA mem with default parameters (version 0.7.12)⁴².
- Twelve wild accessions with higher sequencing depth were downsampled to obtain a similar sequencing depth to the rest of the samples. The alignment bam files were then sorted and PCR duplicates were marked, and a GATK tool HaplotypeCaller was run on each bam file in a GVCF (genomic variant call format) mode. The gVCF files from 440 accessions were consolidated into a single gVCF file, from which single nucleotide polymorphisms (SNPs) and small indels were identified using a joint calling approach.

- Layout figures in **a logical order**
 - Combine the figures **related to the same topic**
 - Write the simple legends
 - Expand into **structured** results
-
- Write **section by section** if possible
 - Write panel by panel if facing some difficulty

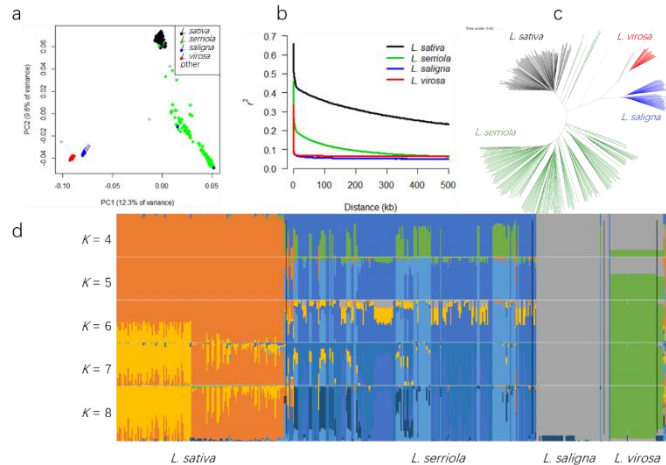
- Make them **self-explanatory**
 - People do not need to read legends to understand
 - Be informative and be pretty
 - State findings in figure titles
 - Explain everything in legends
-
- Learn from good journals
 - Avoid use red & green

Tips for writing result sections

- Each paragraph should have **ONLY one topic sentence**.
- Paragraphs should be **logically organized** around headings or sub-headings.
- Do not count words closely.

Result example: text for figure

- Our phylogenetic result showed that all the *L. sativa* accessions formed a monophyletic clade...
- Model-based clustering analysis revealed **additional** inter- and intra-species relationships.
- The phylogenetic relationships were **also** revealed by the principal component analysis (PCA)
- LD went to supplementary information.



- Start a paragraph to state **the major findings**
 - Discuss in details **topic by topic**
 - Bring out **the significance** in each paragraph
 - End with a **conclusive** paragraph
-
- Do not speculate too much
 - Separate or combine; do not do both

Discussion example: major findings

Discussion

In this work, we analyzed the genome sequences of 445 *Lactuca* accessions representing lettuce crop types and its wild gene pool species. More than 208 million sequence variants were identified, from which we revealed the population structure of the genebank collection and the domestication history of cultivated lettuce.

Discussion example: questions answered

WHY

WHAT

As germplasms of major crops are maintained as genebank collections, understanding the population structure and phylogenetic relationships is of great importance for genebank management and utilization. In lettuce breeding, the GP1 species are used widely as there is no reproductive barrier within the group^{6,7}. Our phylogenetic analyses clarified several issues regarding the taxonomic status of these GP1 species (see the Supplementary Note for a detailed discussion). First, the presumed GP1 species *L. georgica* should be reassigned as it clustered with the GP3 species *L. virosa*. The *L. dregeana* and *L. sagittata* samples are not to be considered as true wild species. Another GP1 species, *L. altaica*, has been considered as conspecific with *L. serriola*^{29,30}, but the plastid phylogeny implied an introgression and fixation of a distantly related plastid haplotype in *L. altaica*. Phylogenetic analyses with additional samples will clarify these taxonomic issues in wild species. Our study also pointed out future directions in germplasm collection and utilization. Among the investigated samples, *L. serriola* from the Caucasus represents the most promising resource because the population from this area showed the highest nucleotide diversity. *L. aculeata* represents another potentially important gene pool, as its phylogenetic position distinct from other GP1 species suggests a different genetic repertoire. Thus, our study provided new insights regarding accession identity and genetic resources for crop improvement, demonstrating the value of whole-genome sequencing in the management of crop collections and the utilization thereof.

HOW

SO WHAT

Personal note

Discussion example: conclusion

Overall, our study constructed phylogenetic relationships within lettuce gene pool species and revealed the genetic basis of human selection during lettuce domestication. The genome sequences and the variation map generated in this study will serve as a valuable resource for lettuce research and breeding in the future.

- Background information
 - Describe **the importance** of your area of study
 - Review the major findings in the field
 - Introduce **the related concepts/pathways/genes**
- Scientific question(s)
 - Raise the question **in a logical way**
 - Stress **its significance**
- Your findings
 - Communicate the major discovery

- The space is limited, normally 1-3 paragraphs.
 - Use only materials present **in the literature**
 - Use only materials **directly related** to your work
- Keep writing
 - Use your background knowledge
 - Do not overthink on details
 - Add (ref/cite) for important information and prepare a reference file

Introduction example: background

Lettuce (*Lactuca sativa* L.) is an important vegetable crop in the Compositae (also known as Asteraceae) family and is widely consumed as salad greens in many countries. Lettuce was first depicted on wall paintings of Egyptian tombs around 2,500 BC^{1,2}, making it one of the oldest known vegetable crops. It is believed that cultivated lettuce was domesticated from its progenitor *L. serriola*, and several hypotheses were proposed regarding the domestication center of lettuce, including Egypt, the Mediterranean area, the Middle East and Southwest Asia¹⁻³. Modern lettuce varieties are classified based on morphological characteristics into leaf lettuce types (namely cos, butterhead, crisp, Latin and cutting) and non-leaf types (stalk and oilseed)⁴. Oilseed, mostly grown in Egypt for seed oil, is considered the most primitive type, while cos lettuce represents the predecessor of leaf types^{2,5}. Despite the morphological variations, different lettuce types share common agronomic traits, such as entire leaf morphology, loss of seed shattering and an absence of spines along the leaf midvein, which are recognized as the domestication syndrome in cultivated lettuce².

- The length is normally two sentences in the second last paragraph.
 - Avoid raising questions in each paragraph
 - Lead directly to your work
- The question is based on the most recent literature.
 - Focus the most related paper
 - Do not ignore any papers

Introduction example: question(s)

Advances in DNA sequencing technology make it feasible to study the genetic architecture in such germplasm collections. A previous RNA sequencing (RNA-seq) study of 240 lettuce accessions demonstrated that different crop types of cultivated lettuce were derived from a single domestication event⁹. However, the domestication history of cultivated lettuce and the genetic basis of human selection remain largely unknown.

- The length is one paragraph.
 - Start with a summary sentence, like “here in study, we carried out a omics approach with ...”
 - Present some details that are enough to understand the scope
 - Bring out the significance in the end

Introduction example: findings

In this study, we sequenced 445 *Lactuca* accessions from 47 countries, comprising the major lettuce crop types and wild relative species. A comprehensive variation map, including 179 million single-nucleotide polymorphisms (SNPs), 30 million insertions/deletions (indels) and 244,866 structural variants, was constructed, from which we analyzed the phylogenetic relationship within the gene pool species and the domestication history of cultivated lettuce. The genetic architecture of domestication traits and introgression regions in resistance clusters were also identified. These sequencing results provide a valuable resource for lettuce research and breeding in the future.

- It should emphasize information on the **MAIN concept/discovery** you want to tell readers.
 - What was **DONE**
 - What was **FOUND**
- It is not easy to find a precise title, but you should have after the manuscript is drafted.



DONE → **Whole-genome resequencing of 445 *Lactuca***
FOUND → **accessions reveals the domestication history of**
cultivated lettuce

PLANT SCIENCE

DONE → **Mutation of a bHLH transcription**
FOUND → **factor allowed almond domestication**

Questions?