PRIMARY RESEARCH   **OPEN ACCESS**

Measuring the Evolutionary Distances between Brassicaceae Species

Bronwen Evans, BCS Student [1], Lingling Jin, PhD [1]

[1] Department of Computing Science, Thompson Rivers University, Kamloops, British Columbia, Canada V2C 0C8

*Corresponding Author: evansb161@mytru.ca

Abstract:
Evolutionary relationships can help us understand the history and evolution of a species. Evolutionary relationships are often discovered or confirmed by molecular phylogeny, which allows us to compare species by their genomes. About 9-15 million years ago, the Brassica genus, which includes canola, broccoli, mustards, and other plants, experienced a whole genome triplication event. This event not only increased the size of the genome, but also the number of duplicated genes within it. We used the Genome REarrangements with Duplications (GREDU) software package to find the double-cut-and-join edit distances between five Brassicaceae species, three of which were from the Brassica genus. GREDU rearranges genomes to find an approximation of the smallest number of rearrangements that must be made to transform one genome into the other. The smaller the number, the closer any two species are predicted to be. GREDU notably supports the comparison of genomes with duplicate genes included in the calculation, which is important when working with Brassica species. We were able to reconstruct the widely accepted phylogenetic tree of the species studied, however we discovered that the GREDU tool may not be best when comparing Brassica species due to the high number of duplicate genes. Future areas of research include determining the actual edit distances between species and analyzing whether GREDU is an appropriate package for conducting comparisons when species have a high number of duplicated genes.

Keywords: genome rearrangements; double-cut-and-join model; Brassicaceae; GREDU; Brassica; evolutionary relationships; edit distance; phylogeny

Introduction

To determine when biological traits first appeared, as well as how long ago species diverged, we study evolutionary relationships. There are many ways to uncover evolutionary relationships, but for about three decades, the prevailing method has been molecular phylogeny. Molecular phylogeny, or determining evolutionary relationships through the use of genome comparison, has grown in use since the 1990s. Typically, molecular phylogeny gives a more accurate picture of evolutionary history than morphological approaches and can be used to solve problems impossible to tackle otherwise [1].

We used the Genome Rearrangements with Duplications (GREDU) software package to find the double-cut-and-join (DCJ) edit distances between five Brassicaceae species. Edit distances are the number of genome rearrangements needed to transform one species into another and are a measure of evolutionary distance. GREDU notably supports comparisons of genomes with duplicate genes included in the calculation, whereas many models ignore duplication [2].

Within the Brassicaceae family is the Brassica genus. Researching these plants is important to Canada because we are the number one producer of canola, a variety of Brassica napus, in the world [3]. Canada grew 20.3 million tonnes of

the crop in 2018 [4]. By exploring the evolutionary history of Brassica, we can better understand one of the vital crops the Canadian economy relies upon.

About 9-15 million years ago, Brassica experienced a whole genome triplication (WGT) event. The WGT event not only led to the speciation of *Brassica rapa, Brassica nigra, and Brassica oleracea*, but it also opened the door for an explosion of variety within these species. Brassica plants now had about three copies of every gene and plenty of room for experimentation. For example, broccoli, cauliflower, brussels sprouts, and kale are all different varieties of the same species, *B. oleracea*. All six cultivated Brassica species show similar versatility and are food staples across the world [5].

We attempted to reconstruct the widely accepted phylogenetic tree of *Brassica rapa, Brassica oleracea, Capsella rubella, Arabidopsis lyrata,* and *Arabidopsis thaliana* using the GREDU DCJ tool to determine whether GREDU could handle calculations involving large amounts of gene duplications.

The rest of the paper is organized as follows: the methods section explains the model used to compute evolutionary distances. Next, the results section shows the edit distances we calculated, followed by the discussion

which evaluates the accuracy of our results. Finally, the conclusion suggests areas of future study.

Methods

We used the GREDU DCJ tool first introduced in [2] to determine edit distances. The DCJ model simplifies the genomes of two species into lists of genes present in both. These genes are connected by adjacencies into sequences, which can be shuffled to rearrange the genome into a new sequence. As the name "double cut and join" suggests, every DCJ operation breaks two adjacencies. The segments created by the cuts are then joined back together in a different way.

The edit distance between any two species is the minimum number of DCJ operations needed to transform one sequence into the other. The longer ago any two species branched apart in their evolutionary history, the more operations will be required, and the higher the edit distance will be. Due to computational complexity, the DCJ tool usually approximates the edit distance. We find meaning in comparing multiple edit distances against each other.

Species data was downloaded from Phytozome. General Feature Format (GFF) files and annotation documents were assembled into raw data files using scripts. These data files only included the genes which were present in both species,

and they were then transformed into DCJ input files by another tool included in the GREDU package.

First, we re-created the original study introducing GREDU using the genomes of mice, rats, and humans from Ensembl. The purpose of the re-creation was to demonstrate the GREDU DCJ tool returned reasonable results in our environment. Then, we performed 10 comparisons between our plant species. All comparisons were completed in the same environment, and all comparisons were limited to a two hour runtime. This means that we took the best answer the tool could find within the time limit given as the result for that comparison. Our aim was to discover whether we could re-create the commonly accepted phylogenetic tree for Brassicaceae and if the edit distances for comparisons involving *B. rapa* or *B. oleracea* were reasonable.

Results

Table 1 shows the edit distances found between mice, rats, and humans using the GREDU DCJ tool. These are how many DCJ operations GREDU determined were necessary to rearrange one genome into another. These results match closely with those published in the original GREDU paper [2].

Table 1. Comparison of the edit distances found by our study and [4].

	Our results	Original study results	Difference
Rat Mouse	670	916	246
Mouse Human	856	894	38
Human Rat	1046	1294	248
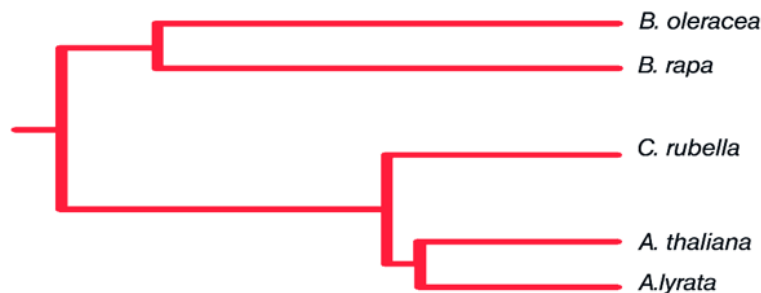
Figure 1. The phylogenetic tree constructed based upon the edit distances in Table 2. The length of the branches represents the distance.

The phylogenetic tree in Figure 1 shows the relationships between Brassicaceae species suggested by the edit distance results we calculated using the DCJ tool. These findings confirm previously studied and commonly accepted phylogenetic relationships. Table 2 shows the edit distances

used to construct this tree. The edit distances involving Brassica species are noticeable larger, especially when compared against the other species, with the greatest edit distance being between *A. lyrata* and *B. oleracea*.

Table 2. Edit distances found by the GREDU DCJ tool.

	B. rapa	B. oleracea	A. thaliana	C. rubella	A. lyrata
B. rapa	0				
B. oleracea	1415	0			
A. thaliana	1689	1752	0		
C. rubella	1740	1693	654	0	
A. lyrata	2088	2130	610	772	0

Discussion

Our re-creation of the original study varied slightly from its results. However, the differences are reasonably accounted for by two main factors. The first is that we were unable to use the same data as was used in the original study because [2] did not include their data versions. We used the most recent genomes available to us at time of calculation, and our updated data likely accounts for much of the discrepancy between our results. The second factor is the environment in which the calculations were run. Differences in the quality and speed of our machines likely affected the edit distances found. For these reasons, we believe that our results are reasonable and demonstrate that our GREDU DCJ tool operated properly.

The phylogenetic tree we were able to construct matched the commonly accepted evolutionary history of the Brassicaceae family. However, the edit distances involving B. rapa and B. oleracea were much larger than expected. One explanation for this could be that the GREDU DCJ tool is unable to accurately calculate the distances between species with as much duplication as Brassica species have. In this case, while the results are a broadly accurate view of Brassicaceae's evolutionary history, GREDU may not be the best tool to use when analyzing the evolutionary relationships of Brassica species. Another possibility is that the DCJ tool simply needs more time to find a smaller, more accurate answer when processing calculations with more duplications. The problem of finding an edit distance in a sequence with duplications has been proven to be NP-hard [2]. It is possible that within the 2 hour time limit we set for all calculations, comparisons involving Brassica species were too complex for a smaller, more accurate edit distance to be reached.

Conclusions

We found that the GREDU DCJ tool could accurately determine the evolutionary history of Brassicaceae species. However, the calculations involving Brassica species returned higher edit distances than expected. GREDU may not be able to find accurate edit distances when Brassica species are involved. For Brassica comparisons, the tool may require more time than it could reasonably be allowed to run. Areas for future study include discovering the actual edit distances between these species, which is to say, finding and confirming the optimal way in which to perform DCJ operations to transform one genome into the other. Furthermore, it is worth investigating whether GREDU is a suitable model for analysing genomes with so much duplication or if a different model should be used.

List of Abbreviations
DCJ: double-cut-and-join
GREDU: Genome REarrangements with Duplications
NP: nondeterministic polynomial time
WGT: Whole genome triplication

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Ethics Approval and/or Participant Consent
This study did not require the approval of an ethics board.

Authors' Contributions
BE: collected data and acquired results, assisted with the analysis of results, and drafted the manuscript
LJ: conceived the study, assisted with the analysis of results, and gave final approval of the version to be published

References
[1] Brown T. Chapter 16. Genomes [Internet]. 2nd ed. Oxford: Wiley-Liss; 2002 [cited 10 May 2019]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21122/
[2] Mingfu Shao, Yu Lin, and Bernard M.E. Moret. An Exact Algorithm to Compute the Double-Cut-and-Join Distance for Genomes with Duplicate Genes. Journal

of Computational Biology. 2015 May;22(5):425-35. https://doi.org/10.1089/cmb.2014.0096

[3] Leading producing countries of rapeseed in 2018/2019 [Internet]. Statista. 2019 [cited 10 May 2019]. Available from: https://www.statista.com/statistics/263930/worldwide-production-of-rapeseed-by-country/

[4] Statistics Canada. Production of principal field crops, November 2018. Statistics Canada; 2019 p. 5.

[5] Cheng F, Wu J, Wang X. Genome triplication drove the diversification of Brassica plants. Horticulture Research. 2014;1(1). https://doi.org/10.1038/hortres.2014.24

Citation
Please cite this article as follows:
Evans B, Jin L. Measuring the evolutionary distances between Brassicaceae species. URNCST Journal. 2019 Oct 09: 3(9).
https://urncst.com/index.php/urncst/article/view/156
DOI Link: https://doi.org/10.26685/urncst.156

Do you research in earnest? Submit your next undergraduate research article to the URNCST Journal!
| Open Access | Peer-Reviewed | Rapid Turnaround Time | International |
| Broad and Multidisciplinary | Indexed | Innovative | Social Media Promoted |
Pre-submission inquiries? Send us an email at info@urncst.com | Facebook, Twitter and LinkedIn: @URNCST
Submit YOUR manuscript today at https://www.urncst.com!