Langara.
THE COLLEGE OF HIGHER LEARNING.

**Research and Scholarly Activity Fund Award Final Report**

Name:

Department:

Faculty:

In no less than 350 words, please reflect on the status of the project for which you received an RSAF award. You *may* wish to address: potential new avenues for the project enabled by the award; how the funded project builds on previous projects and/or collaborations; any publications, conferences, talks, classroom modules, etc., coming out of the funded project; the outcomes of collaboration facilitated by the award; the potential for future research and scholarly activity projects made possible by the award; or the funded project's impact on your activities in general at the College. As this report will be posted on the Scholarly Activity website, please submit an electronic copy.

The RSAF award final reports may also be published in the Research and Scholarly Activity annual report. You will be notified of this and given final approval of content before publication.

You may also be asked to participate in events promoting research and scholarly activity at the College.

Please write in the space below or attach an additional PDF page.

---

## ULTRA DNA BARCODING FOR IDENTIFICATION OF HOP VARIETIES

**JI YONG YANG**

## BACKGROUND

Hops (*Humulus lupulus*) are increasingly becoming an important cash crop in BC primarily used to give beer a characteristic aroma and bitterness. There are over 100 different varieties of hops, and each can impart a different beer flavour profile. In recent years, beers with unique combinations of hop varieties, mainly India pale ales or IPAs, have markedly increased in popularity and have been a cornerstone of BC's burgeoning craft brewing industry. However, distinguishing them based on morphological characters alone can be difficult. An efficient and reliable DNA marker-based genotyping method is needed. Chloroplast DNA is ideal for identifying hop varieties because it is maternally inherited, and the hops are the female flowers of the hop plant. Our goal is to develop a DNA barcoding system using whole chloroplast genomes (i.e. ultra-barcoding) for identification of hop varieties.

## METHODS

We sampled and extracted DNA from 35 hop varieties. We chose DNA from ten popular hop varieties for genome sequencing. We used next-generation sequencing (NGS) to assemble the whole chloroplast genomes for ultra-barcoding of the ten hop varieties. Starting with the NGS data, seven different bioinformatics software were used (see Fig. 1). First, the NGS data reads

were trimmed using the software Trimmomatic (1) and then aligned using the software Burrows-Wheeler Aligner (BWA (2)). We used two software packages, Binary Call Format (BCFtools (3)), and Genome Analysis Toolkit (GATK (4)) for variant and genotyping calling of the ten hop varieties. The result Variant Call Format (VCF) files were used to construct phylogenies using the software (MEGA7 (5)). We also used the software NOVOplasty (6) and DOGMA (7) to assemble and annotate the chloroplast genomes, and the software cpGAVAS (8) to construct the circular map of the chloroplast genome.

## RESULTS AND DISCUSSIONS

The hop's chloroplast genome (Columbus variety) is 153,547 bp, circular, and contains 129 genes (Fig. 2). The whole chloroplast genomes from the ten varieties were compared and used to construct phylogenetic trees. The following three types of phylogenies were performed: A) Maximum Likelihood (9), B) Neighbour Joining (10), and C Minimum Evolution (11) (Fig. 3). All three phylogenies revealed very similar evolutionary histories. They all showed two distinct clades of hop varieties, and based on the long branch lengths the two groups diverged some time ago. The phylogenies also show that each hop variety occupies a unique branch tip suggesting that they differ from one another. We demonstrated that by sequencing the entire chloroplast genome, we were able to find sufficient DNA variations to identify each of the ten different hop varieties uniquely. Our next major goal will be to find a set of unique markers (i.e. SNPS, SSRs, Indels) to make the identification of hop varieties faster and easier. We identified particular regions in the chloroplast genome that appear to be highly variable. We have sequenced these regions on all of the ten hop varieties, and we are in the process of finding unique SNPs, and polymorphic SSRs and indels.

## STUDENT PARTICIPATION AND ACKNOWLEDGEMENT

I would like to thank the following SWAP students: Gopala Bhagavatula and Adam Todd for their assistance in DNA extractions and amplifications, and Andreii Ponomarev for his valuable aid in bioinformatics applications. I would like to thank Langara's Biology department for providing me with over 100 SWAP hours for these students. This study was funded by the Langara Research and Scholarly Activity Fund (RSAF).

## FUTURE STUDY

I have plans for a research project involving 'wild' hops in BC. Hop farming was big in BC back in the late 1800s and early to mid-1900s. The industry dried up, but some of their hops escaped into the wild. It would be interesting and to determine the identity of these 'wild' hops. It could even be possible, due to over 100 years of evolution (hybridizations and mutations), that some of the 'wild' hops might be genetically unique to BC. I have been in contact with two potential industry partners, Sam Quinlan from Harvesters of Organic Hops, and Jody Hammell from BigRock Brewery to investigate the genetics of BC's 'wild' hops.

## INTERVIEWS, PRESENTATIONS AND PAPER

I was interviewed by Langara Voice about my hop study. The article, "DNA coding system hopes to help local craft beer industry", was published on March 23, 2017. I presented a

research poster at the Langara's Scholarship Cafe 2017 event. I also gave a talk at the Beaty Biodiversity Museum at UBC. The event was titled "Beaty Nocturnal Hops". I also plan to submit a paper on "Ultra DNA barcoding in hops" for publication at a top field journal.

**COLLABORATIONS**

I have been in contact with Sam Quinlan from Harvesters of Organic Hops, and Jody Hammell from BigRock Brewery to investigate the genetics of BC's 'wild' hops (see Future Study section). We are talking about applying for an NSERC Engage Grant.

**REFERENCES**

1) Bolger A M, Lohse M, and Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30: 2114-20.

2) Li H and Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics 26: 589-595.

3) Danecek P, Auton A, Abecasis G, et al. (2011) The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

4) DePristo MA, Banks E, Poplin R, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

5) Sudhir K, Stecher G, and Tamura K (2015) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. Molecular Biology and Evolution 33: 1870-1874.

6) Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data Nucleic Acids Research kw 955.

7) Stacia K, Wyman RK, Jansen R, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA Bioinformatics, 20: 3252–3255.

8) Chang Liu, Linchun Shi, Yingjie Zhu, et al. (2012) An integrated web server for the annotation, visualisation, analysis, and GenBank submission of completely sequenced chloroplast genome sequences BMC Genomics, 13: 715.

9) Tamura K and Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10: 512-526.

10) Saitou N and Nei M (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

11) Rzhetsky A, and Nei M. (1992) A simple method for estimating and testing minimum evolution trees. Molecular Biology and Evolution 9: 945-967.

12) Felsenstein J (1985) Confidence limits on phylogenies: An approach using the boostrap. *Evolution* 39: 783-791.

**FIGURES**

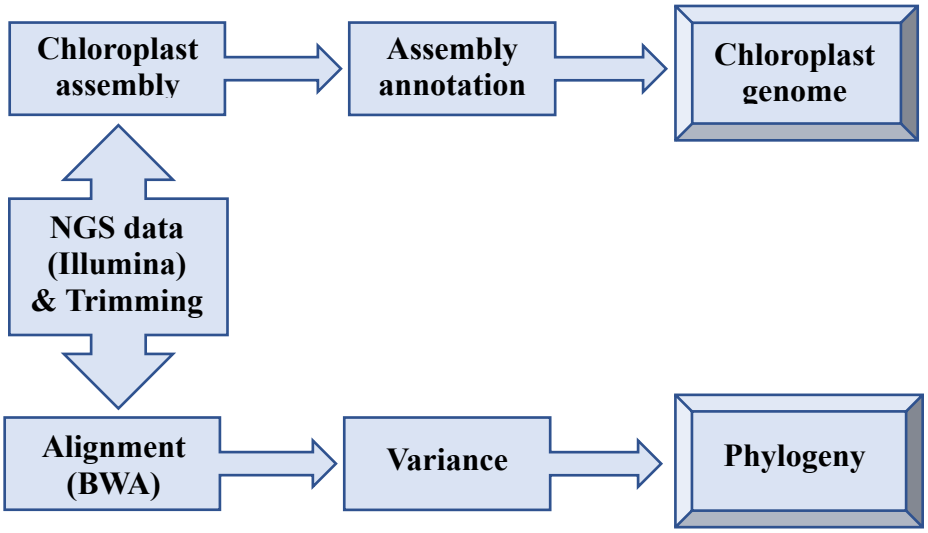


Figure 1. Bioinformatics flow chart showing the step by step procedures to produce a chloroplast genome map and a chloroplast genome phylogeny. Starting with the NGS (next-generation sequencing) data, seven different Bioinformatics' software were used. The terms in brackets represent the names of the methods or software used to carry out those steps.

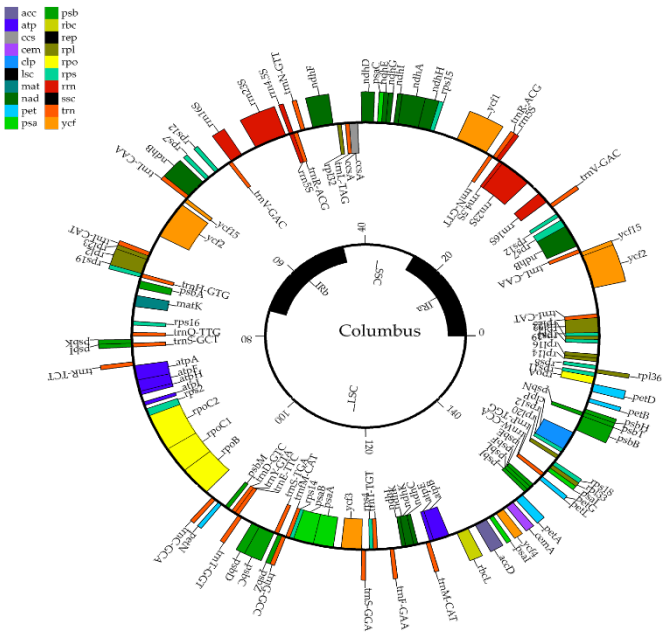


Figure 2. Gene map of the hop's chloroplast genome from the Columbus variety. The genome size is 153, 547 bp, and it consists of 129 genes. The different colour bars represent the different genes, and the black inner bars the inverted repeats.

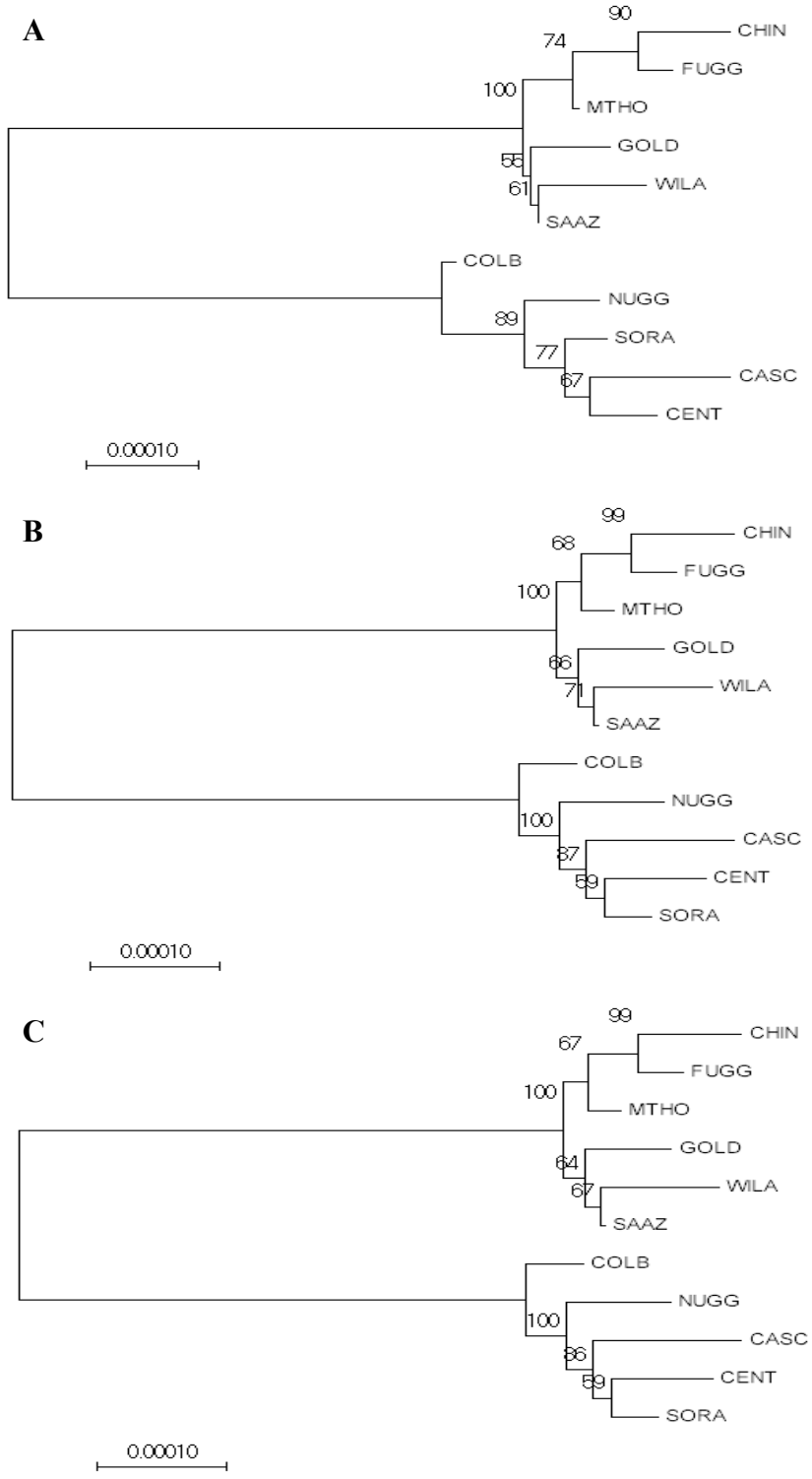


Figure 3. The phylogeny trees of the ten hop varieties were inferred by using three evolutionary distance methods: A) Maximum Likelihood (1), B) Neighbour Joining (2), and C) Minimum Evolution. In all three phylogenies the percentage of replicate trees in which the assigned taxa grouped together in the bootstrap (1000 replicates (12)) are shown next to the branches.