

Comparing the genetics of Black Madeira KK, Figo Preto and Craven's Craving

I am nervous to post this because I am afraid of the storm.

But let the storm commence!

Some background:

I have a Masters in Science in Biochemistry and have been involved in genetics for over a decade. I have first hand experience conducting research using PCR and sequencing methods.

I work for a company that provides sequencing services. The sequencing services provided for these fig samples were for training and validation of a new instrument into our fleet. I was lucky to be in the right place and right time to help provide some unique samples. I did not have to pay for these samples to be sequenced. I doubt I will be in a similar position in the future. If you want more sequencing completed, it is \$1000+ a sample.

Whole Genome Sequencing (WGS) vs PCR: A few members have posted some results from genetic testing for figs via PCR. PCR is great tool but very different from WGS. PCR does not analyze the whole genome. In PCR, targets are chosen and DNA primers are designed to amplify those selected regions of DNA. During PCR, millions copies of the selected targets are generated, they range from a few hundred base pair to a couple thousand. The PCR results are then ran on a gel and target is visualized with a stain. In theory different fig varieties could produce targets of different sizes and when a few targets are looked at together, one could differentiate different varieties. This is a common technique used in agriculture, some species of plants have well qualified targets and are reliable, ie corn or strawberries. However, from my research there doesn't seem to be much consensus on targets for figs with researchers around the globe are still exploring better targets. This PCR technique was also developed for commercial breeding of new plant varieties. Breeders will sprout thousands of offspring from a cross and use this technique to cut offspring similar to the parents. The goal of the breeding is to discover unique offspring, not ones similar to the parents. If you can cut similar offspring, then you can save a lot of time, space and money. Using this PCR technique and backing it into the application of determining clones or lineage, in MY OPINION probably isn't going to be reliable. My opinion on this comes from reading many papers of researchers trying to find reliable PCR targets for figs. New targets are explored but fail in the researchers own tests. Research is ongoing and no consensus has seem to be made. If someone has come across papers that says otherwise, I have happy to read it. I have also yet to read a paper that confirms their PCR with WGS results.

Another option for genetic testing for figs, would be to find a "finger print" region of DNA that could be targeted for PCR amplification and then conduct sanger sequencing. However, targeted next gen sequencing platforms have become more economical then sanger sequencing in recent years. However, I have yet to come across any literature that provides a "finger print" region for figs. This would be the quickest and most cost effective method for determining clones or lineage.

Wole Genome Sequencing (WGS) is the best method for determining clones or relationships. A PCR method does not sequence the sample DNA. The first step in WGS extraction of the DNA from the plant material. For this experiment, I put on a pair of gloves, a broke off a dime sized piece of leaf and changed gloves between sample collection. The collected samples were placed in a tube

with DNA stabilization media to inhibit and DNA degradation. DNA was extracted with a an off the shelf DNA extraction kit for plants. The extracted DNA was check for quantity and quality. The purified DNA then was used to generate a library. The prepared sample was sequenced on an Illumina NovaSeq X. The data was sent to a lab with an High Performance Computer (HPC) Cluster to be demultiplexed, rebuilt and aligned to the published Ficus carica reference genome. Next a data scientist using the HPC compared the whole genomes to each other using a standardized bioinformatic algorithm to determine lineage. As one can imagine this whole process is expensive and time consuming. Since this wasn't an official study and I had to collect my samples in a short time, I wasn't able to get a better collection of samples. I knew that I wanted to investigate the lineage of Black Maderia, Figo Preto and Craven's Craving, that is why those were chosen and to provide and as an analogous comparison, three separate leaves were sampled from a single unknown male seedling outside one of our buildings. I have grown all these trees out and have sampled fruit. All seem to be true to type.

The figs sent in for sequencing:

Zaffiro – Profig (ZF)

White Madeira #1 – Harvey (WM)

Unk Pastiliere – Laguna Hill Nursery, provided by well-known OC collector (UP)

Olympian – Harvey (OLY)

I-258 – Harvey (I258)

Black Maderia KK – Harvey (BM)

Figo Preto - Laguna Hill Nursery, provided by well-known OC collector (FP)

Craven's Craving - Laguna Hill Nursery, provided by well-known OC collector (CC)

Murphy – Unknown male seedling outside one of our buildings.

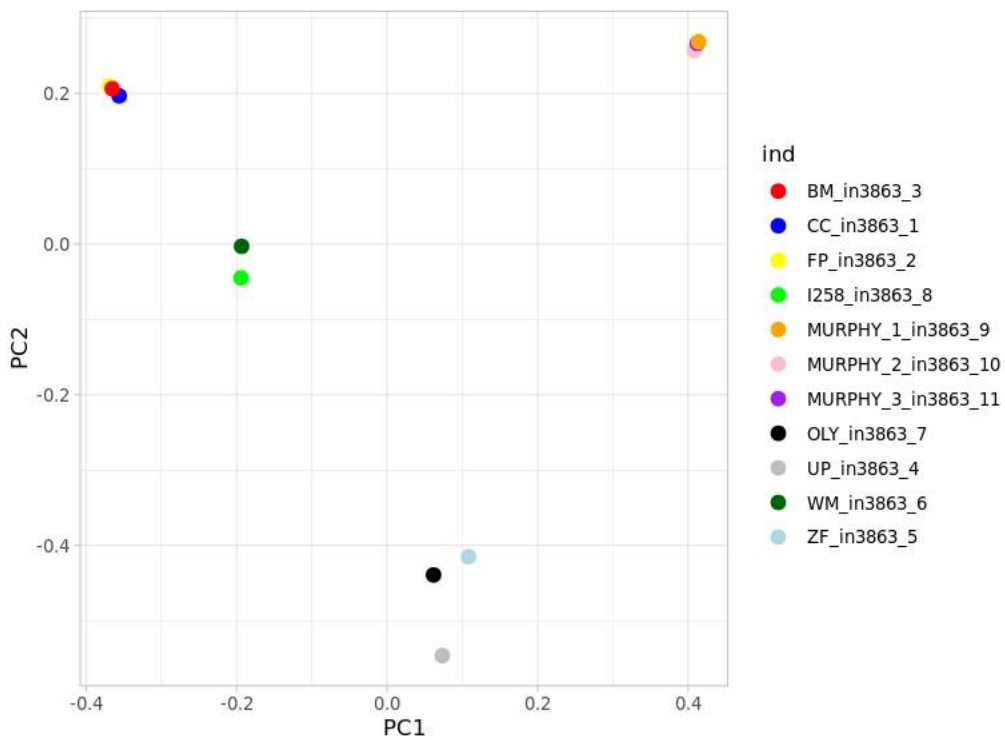
Now it is time for the results. As I explained earlier, experimental design is not perfect. I am well aware. This does provide solid evidence of the relatedness of the fig varieties used in this study. I also plan on using other bioinformatic algorithms to validate these results. However, if I say that BM, CC and FP are clones from a mother tree this would be analogous to grabbing three unique samples of the same tree. When the analysis is complete and we calculate the similarity of the known clone and compare them to BM, CC and FP. If the values show a similar value, I think it is fair to say that BM, CC and FP are clones.

Murphy 1,2 and 3 = .86

BM, CC and FP = .85

These results have very high correlation and it is safe to say that BM, FP and CC are propagation off the same mother tree.

For reference here are the other values:



Murphy to CC = 0

Murphy to FP = 0

Murphy to BM = 0

Murphy to UP = .147

Murphy to Zaffiro = .235

Murphy to WM#1 = 0

Murphy to Olympian = .173

Murphy to I258 = .0

CC to UP = .118

CC to ZF = .140

CC to WN = .478

CC to OLY = .122

CC to I258 = .476

FP to UP = .132

FP to ZF = .0

FP to WM = .488

FP to OLY = .113

BM to UP = .131

BM to ZF = .147

BM to WM = .488

BM to OLY = .114

BM to I258 = .476

UP to ZF = .297

UP to WM = .488

UP to OLY = .216

UP to I258 = .136

ZF to WM = .168

ZF to OLY = .222

ZF to I258 = .194

WM to OLY = .172

WM to I258 = .329

OLY to I258 = .210

The rest of the results seem to support the BM, FP and CC evidence. The scoring is able to separate out all of the varieties. For those of you who familiar with the linages of these varieties, may see some correlation. The scoring can be used with below to determine relatedness.

- **0.00 to 0.125:** Unrelated
- **0.125 to 0.25:** Possible third-degree relatives (e.g., first cousins)
- **0.25 to 0.5:** Possible second-degree relatives (e.g., half-siblings)
- **0.5:** Possible first-degree relatives (e.g., siblings or parent-offspring)

The algorithm used for the analysis was developed for correlation relationships but did not include any twins in the study. A value for twins was not described, However, for this study, proving three samples off the same tree, shows that .85 value could be used to set as twins/clones.

The other varieties relatedness when compared to BM, FP and CC also support clone theory.

UP to CC, FP and BM = .118, .132 and .131

ZF to CC, FP and BM = .142, .0 and .147

Comparing all the other varieties, show a similar trend

Comparing the Murphy 1, 2 and 3 samples to ZF. .224, .235 and .226

Cross referencing shows a similar pattern and supports the BM, CC and FP clone theory.

In conclusion it is safe to say that CC, FP and BM are clones. I am sure not very many people are surprise by this but some maybe are. A member of the OurFigs forum discovered the Craven's Craving tree and can attest to his unique discovery. I do not discount that and I am not trying to undermine the discovery. From my experience growing CC, FP and BM, I was confident that CC would not be a clone to FP and BM. However, the sequencing results have shown otherwise.

As I stated earlier, I will continue to analyze these samples, but I doubt I will see any difference in results. Also, if I get the opportunity to process more figs, I can add them to the database and rerun the analysis. If anyone wants a copy of the WGS sequences, I am happy to share if you have the capacity to analyze the sequences. I will also warn you that you should NOT download this data just for fun. It is multiple TBs of data. If you don't have a large hard drive or fast internet connection, you are wasting your time. Also, data on some storage media can degrade overtime. Hashing this much data to ensure a complete download will probably break your computer. Also to analyze the data, you need access to a HPC cluster or spend \$1000s in AWS or Azure. What I am trying to say, it that is data something the average person can hold on to or process

Thanks!

Chris

I forgot to add the acknowledgements! I did not do this myself. My main effort was collecting the samples, being lucky and pulling a lot of favors. Thank you to the Services Director, the scientists who extracted the DNA and did the library prep, the scientists who ran the samples on the instrument and the bioinformatician who conducted the analysis. A lot of hard work went into this. I am very grateful.