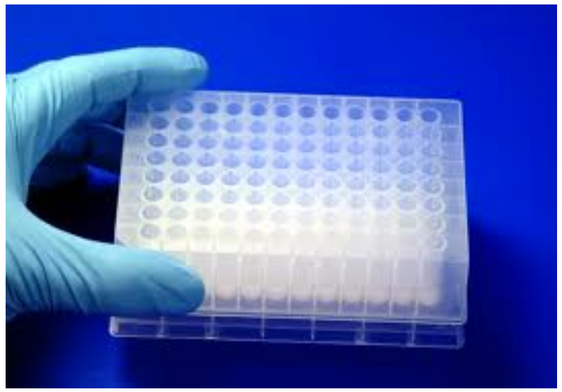
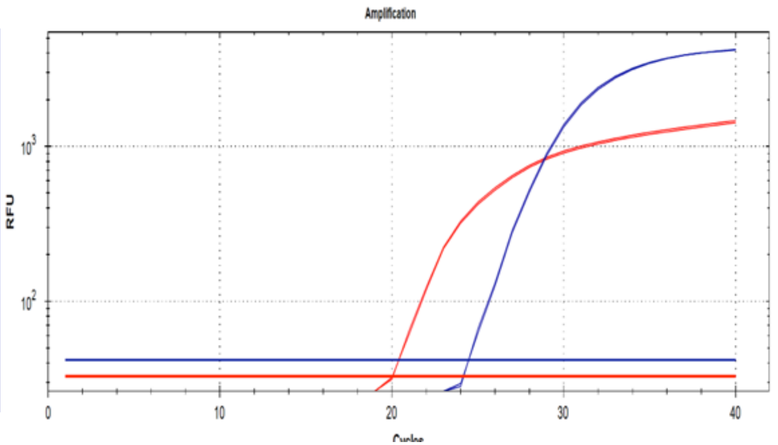
Conflict of Interest Statement

Speaker is an employee of
Medicinal Genomics & Courtagen Life Sciences

MEDICINAL GENOMICS





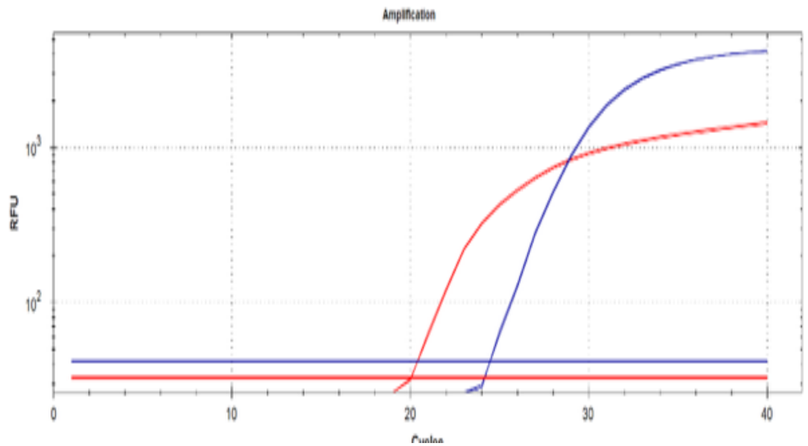
ICRS 2015

SEQUENCING OF THREE MALE CANNABIS GENOMES AND DEVELOPMENT OF MULTIPLEX QPCR ASSAYS FOR RAPID MALE SEX DETERMINATION

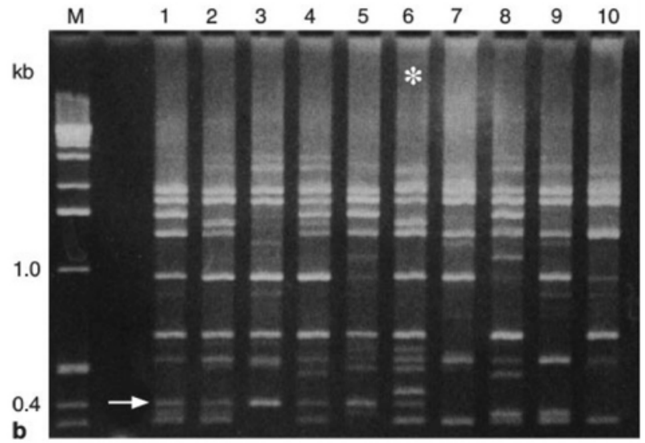
Sex Determination

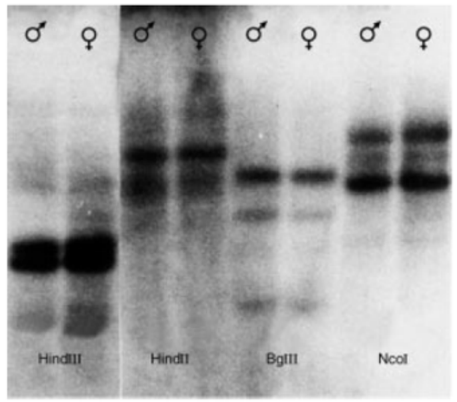
qPCR of Medicinal Genomics discovered Y Chromosome specific sequences. Conventional Mandolino primers are not Y specific. They target a highly polymorphic repeat called MADC2.

Medicinal Genomics



Mandolino





MEDICINAL GENOMICS

A Closer look at MADC2 (Mandolino)

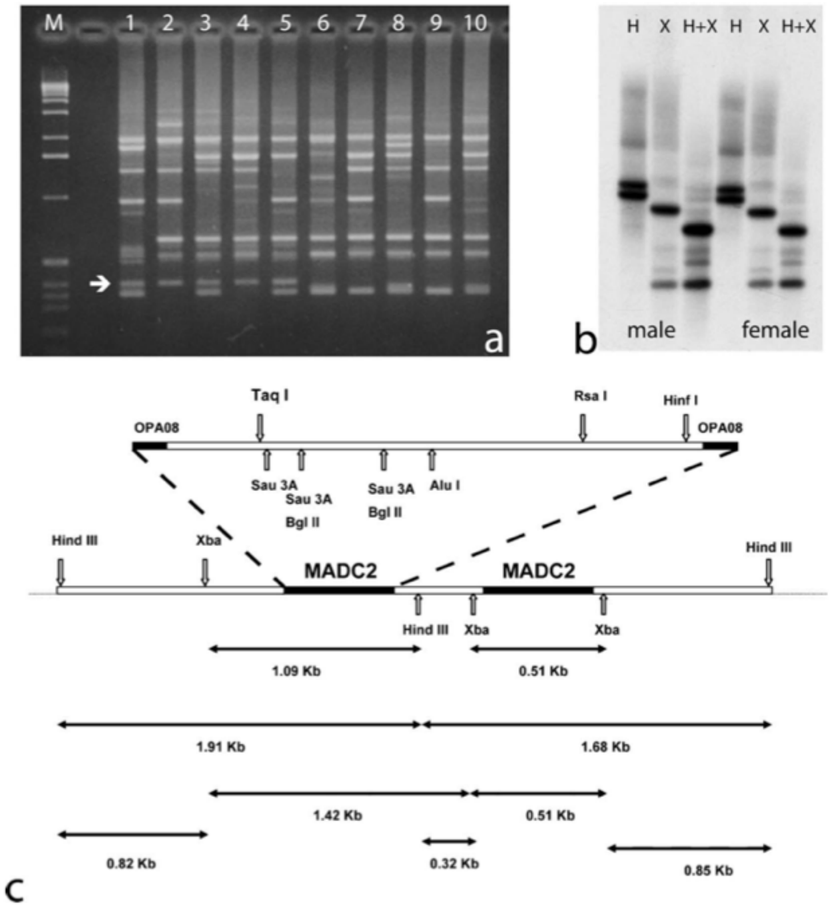


Figure 2. (a) RAPD pattern generated by the primer OPA08 (Operon Technologies, USA). The arrow indicate the 400 bp male-associated marker, present in the male plants (lanes 1–5) and absent in the female plants (lanes 6–10). M, molecular weight standard. (b) Hybridization pattern obtained using the cloned 400 bp RAPD fragment visible in a) on digested DNA from male and female plants restricted with *Hind*III (H), *Xho*I (X) and with both enzymes (H + X). Note the absence of male–female polymorphism. (c) Restriction map of the genome region containing the sequence corresponding to MADC2.

AF364955, MADC5 and MADC6, by Torjek et al.), and also female-associated RAPD markers were reported (Shao et al., 2003), though in the latter case no chromosome-specific location can be hypothesized, and it was not specified whether this marker was able to distinguish female from monoecious plants.

The monoecious trait is another very important character for which MAS would be extremely useful.

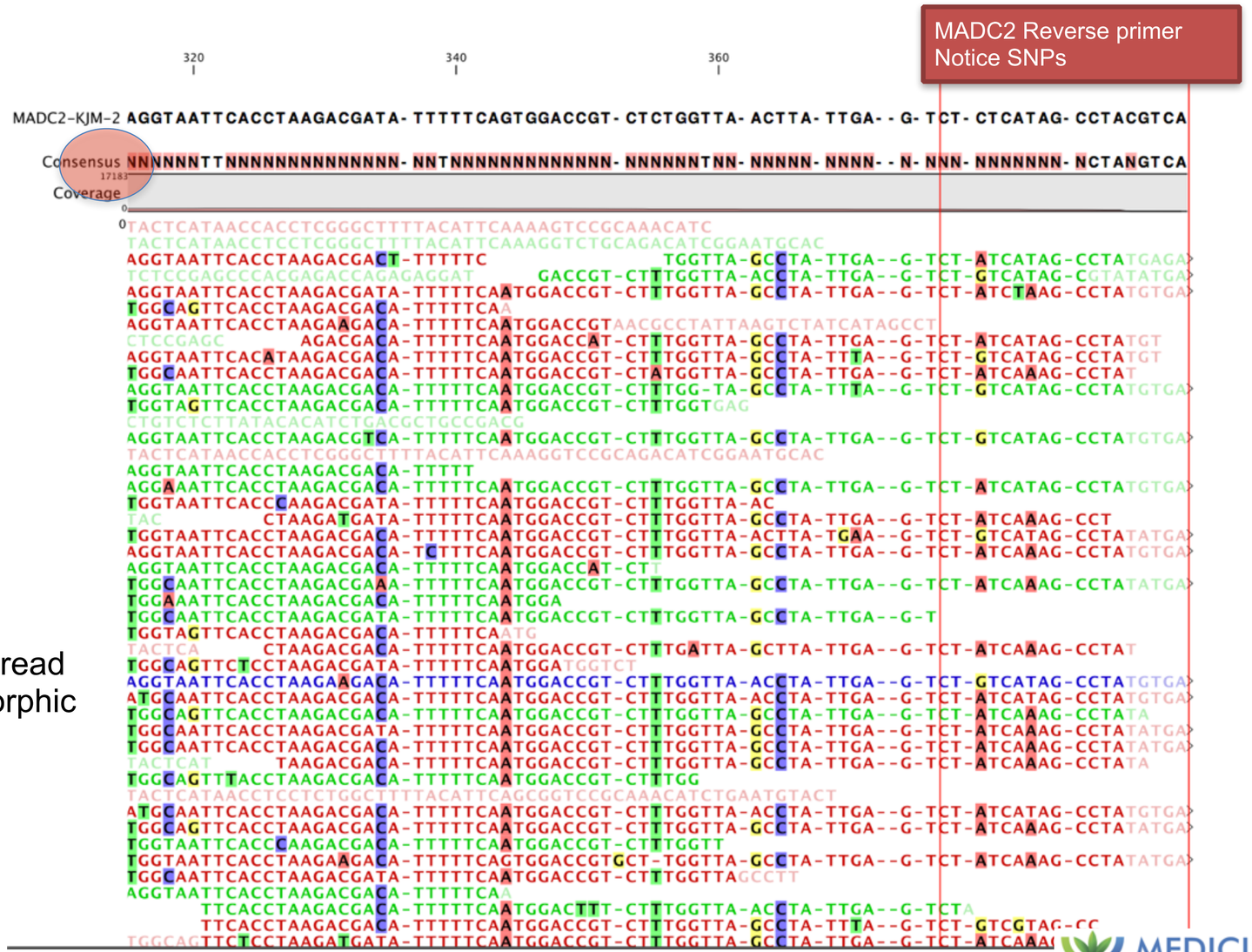
Mandolino primers target MADC2 and generate a smear on a gel.

Notice MADC2 is a repeat and only has a pre-genome restriction enzyme map.

MEDICINAL GENOMICS

A NextGen view of MADC2 is complex

Sequencing Alignment of WIFI male reads to MADC2 shows 17,000X coverage in a ~30X coverage genome.
A repeat with high probability of producing strain to strain variance.
Not ideal for a qPCR target as all strains generate products.



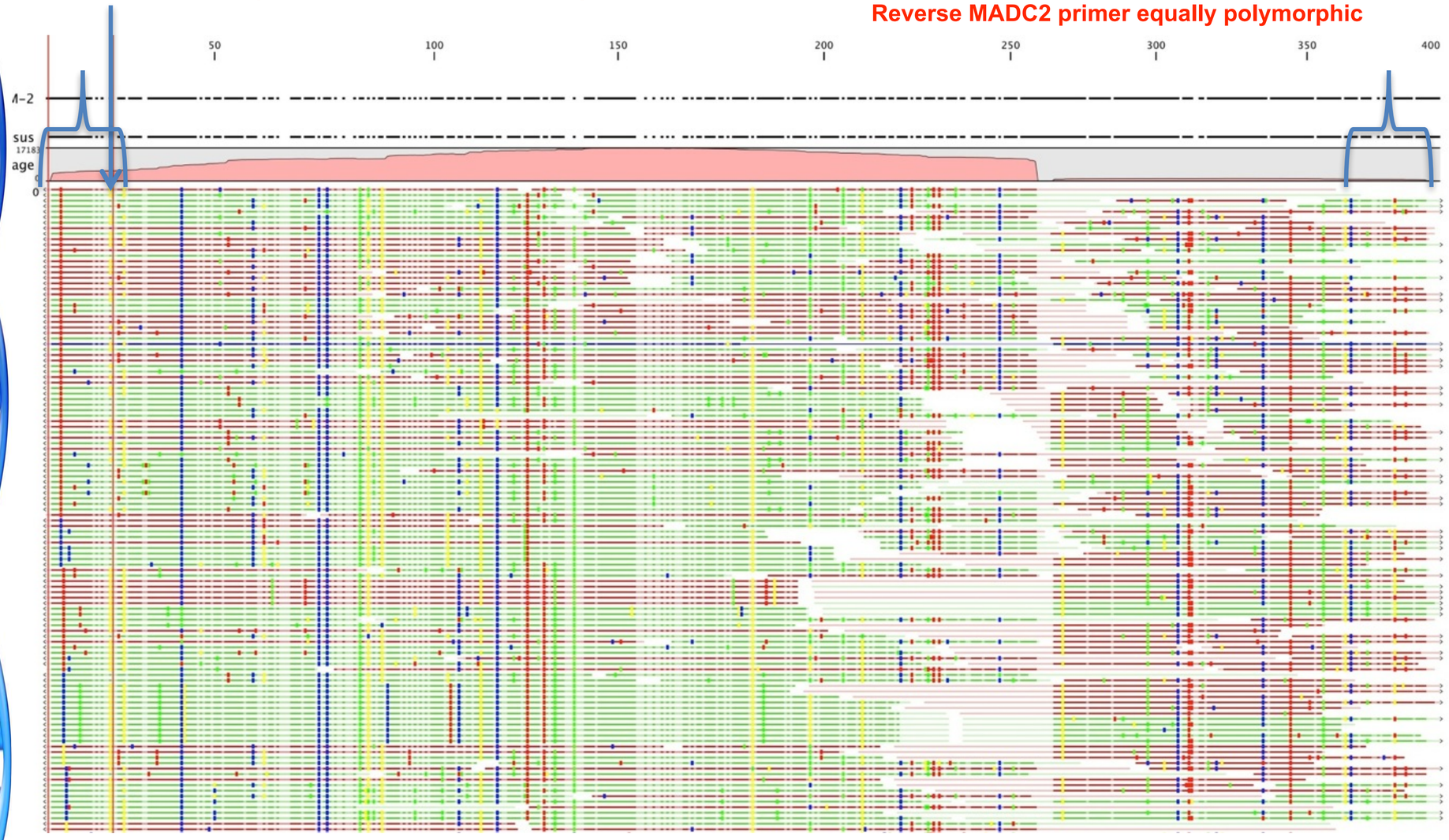
Only 1 paired read
Highly Polymorphic loci

MEDICINAL GENOMICS

Mandolino Primers sit on SNPs

MADC2 Forward primer sits on 3’ SNPS in a repeat!

Reverse MADC2 primer equally polymorphic

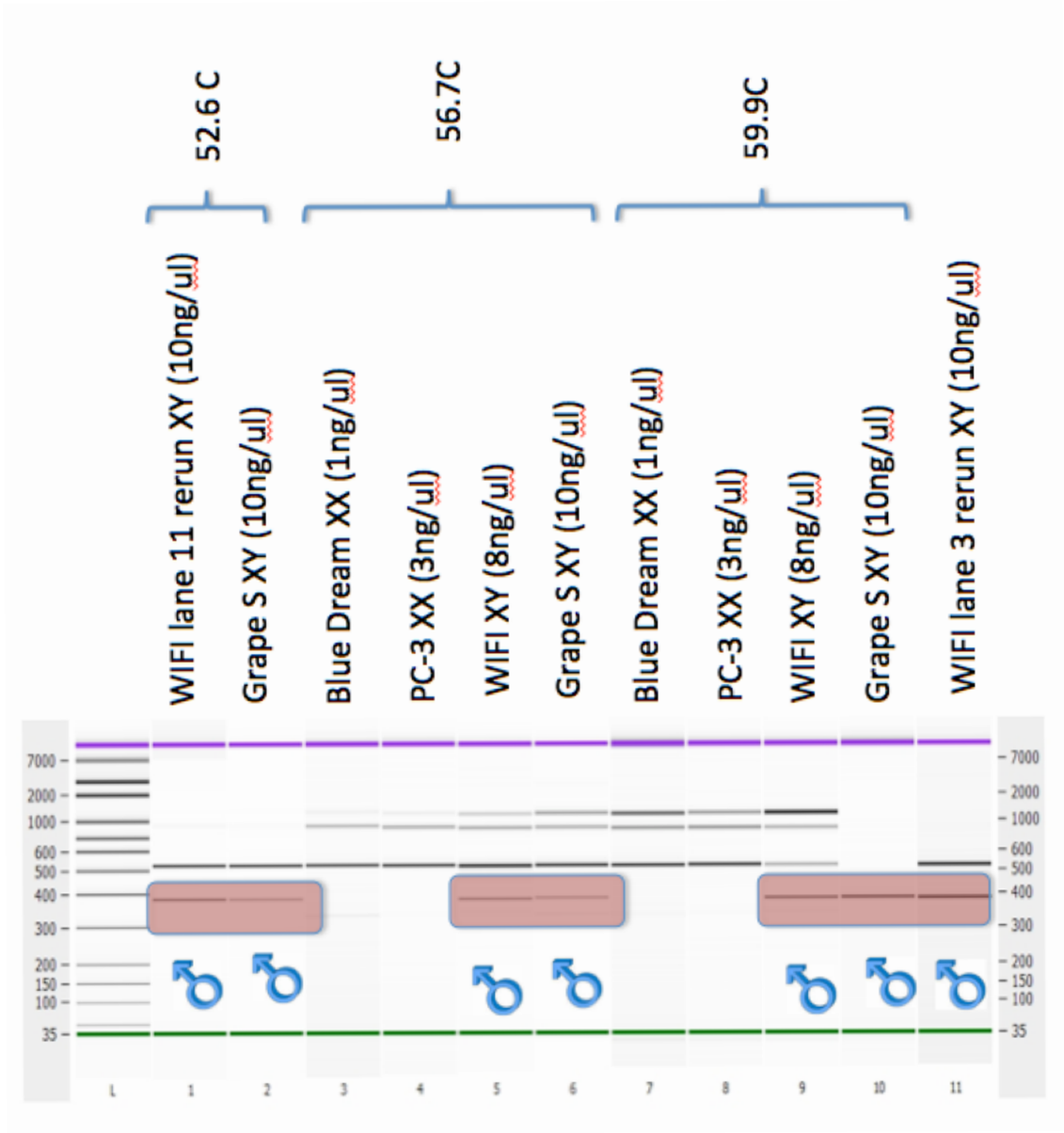


This can be substantially improved with the sequence of a few males

Horizontal Red and Green lines are Forward and reverse reads respectively.
Vertical lines are SNPs. Blue reads are paired reads of the correct length (only1)

MEDICINAL GENOMICS

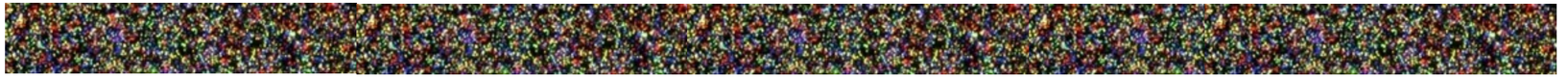
Perhaps Tm Optimizations for MADCx?



Increasing Tm = More bands not less?

MEDICINAL GENOMICS

Solution to Sex Determination



Sequence truck loads of plants to identify better Y chromosome markers

12 different Female (XX) cannabis strains were sequenced and assembled
2 different Male (XY) cannabis strains were sequenced and assembled

- **Contigs were filtered based on**
- **1)No microbes**
- **2)No repeats**
- **3)Limited polymorphism rate between 2 males/4 parents**
- **4)GC content and good amplification design criteria with Primer3**

Taqman assays were designed and tested on males and females and products evaluated on agilent microchip analysis.

The MADC2 regions described by Mandolino were also evaluated on Agilent and by read mapping analysis of all Male Whole Genome Sequence (WGS) alignment to the contigs in female and male assemblies that match those primer sequences (very interesting results).

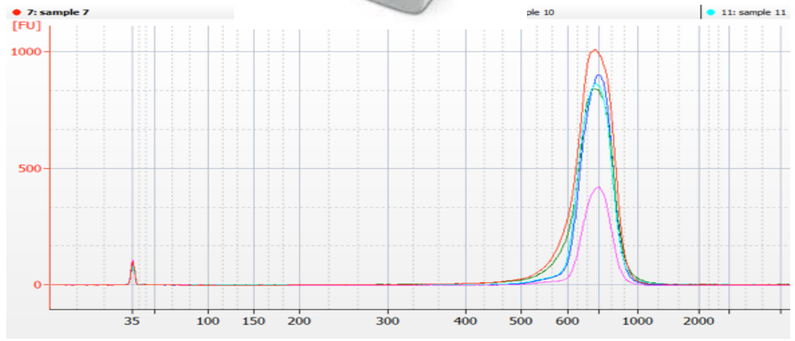
New assays designed to avoid the repetitive nature of MADC2.

MEDICINAL GENOMICS

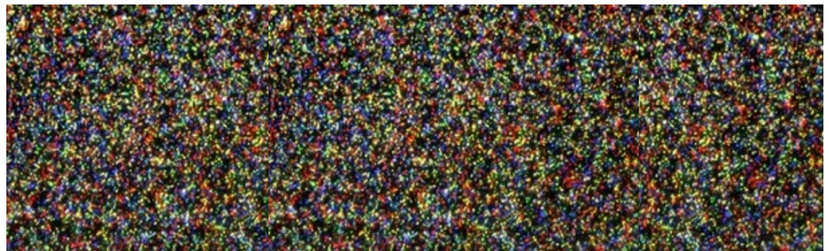
Experimental Design

Whole Genome Shotgun 15 Cannabis Genomes
2 Male and 12 Female

Size Select 800bp libraries for 2x250bp sequencing



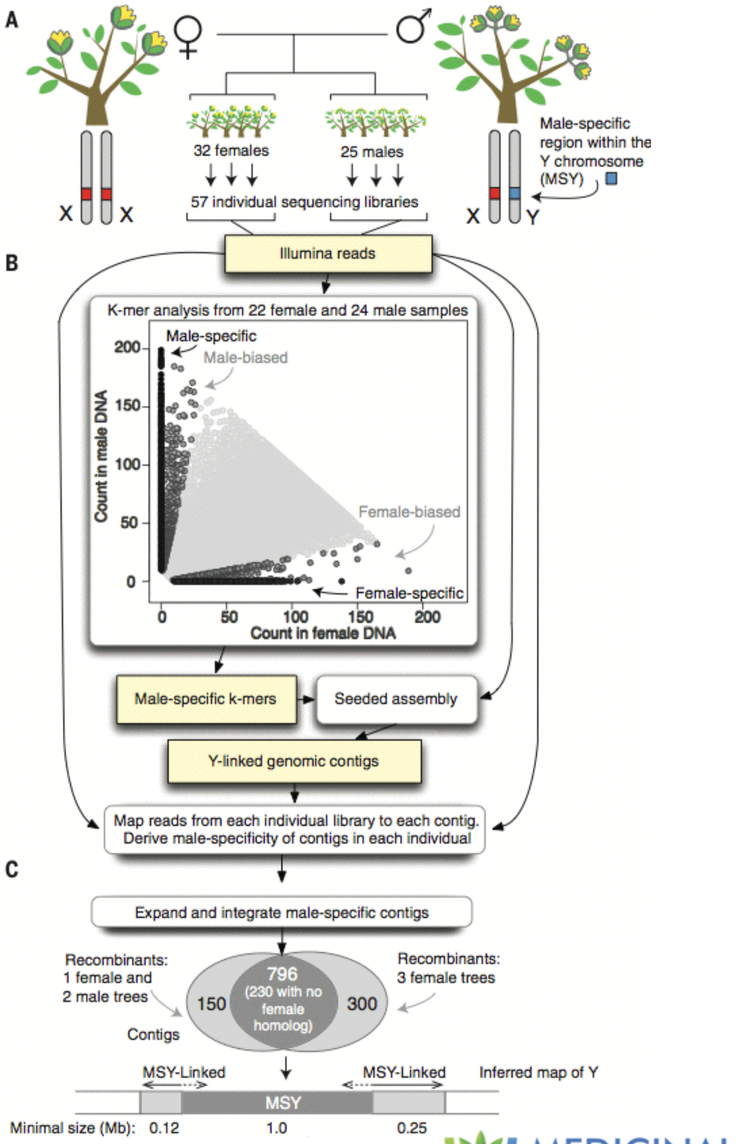
ILMN Sequencing 50-100 Million paired 250bp reads/strain



MEDICINAL GENOMICS

in silico genome subtractions
Female vs Male genome assemblies

- Modified Akagi *et al.* approach.
- Smaller Cannabis genomes and previous assemblies of cannabis can inform a more efficient approach.
- 80M reads per Male with 2x250bp reads enables assembly of 300Mb N50/800Mb complete drafted genomes where 1000s of >500bp contigs are Male specific.
- Filter and screen top candidate contigs for qPCR assays



MEDICINAL GENOMICS

Assembly statistics

WIFI:
2 DNA Extractions
Total Reads: 98,076,120
Mapped Reads : 94,409,402 (99.60% & 96.26%)
Contigs from unmapped reads : 636
Contigs that blast to Bacteria : 18
Contigs that blast to other Eukaryotes : 36

Grape Stomper:

Total Reads : 88,234,364
Mapped Reads : 87,864,458 (99.58%)
Contigs from unmapped reads : 1166
Contigs that blast to Bacteria : 195
Contigs that blast to other Eukaryotes : 67

Australian Bastard
Total Reads : 80,358,484
Mapped Reads : 75,593,886 (94.07%)
Contigs from unmapped reads : 8545
Contigs that blast to Bacteria : 7278
Contigs that blast to other Eukaryotes : 132

2326 unique contigs combining all the contigs from the 3 strains.

No Putative Y contig BLAST Hits to female Hops genome

				match length	query start	query end
ctg7180000009773	gi|147786918|emb|AM423511.2|	Vitis vinifera	Eukaryota	3044	1	3023
ctg7180000005237	gi|147786918|emb|AM423511.2|	Vitis vinifera	Eukaryota	1855	130	1956
ctg7180000009722	gi|147792470|emb|AM462571.2|	Vitis vinifera	Eukaryota	1846	14	1828
ctg7180000009772	gi|147786918|emb|AM423511.2|	Vitis vinifera	Eukaryota	1392	1	1387
ctg7180000009847	gi|147863205|emb|AM437310.2|	Vitis vinifera	Eukaryota	1125	1	1101
ctg7180000009786	gi|147836563|emb|AM482474.2|	Vitis vinifera	Eukaryota	606	229	833
ctg7180000005881	gi|11544478|emb|AL353701.15|	Homo sapiens	Eukaryota	344	1	344
ctg7180000005229	gi|703135881|ref|XM_010107707.1|	Morus notabilis	Eukaryota	336	1117	1451
ctg7180000005209	gi|388507803|gb|BT142174.1|	Lotus japonicus	Eukaryota	320	126	444
ctg7180000005987	gi|568839835|ref|XM_006473819.1|	Citrus sinensis	Eukaryota	283	381	662
ctg7180000009846	gi|208609064|dbj|AB430329.1|	Lotus japonicus	Eukaryota	277	1070	1344
ctg7180000005396	gi|24899167|dbj|AB095909.1|	Vicia faba	Eukaryota	232	353	583

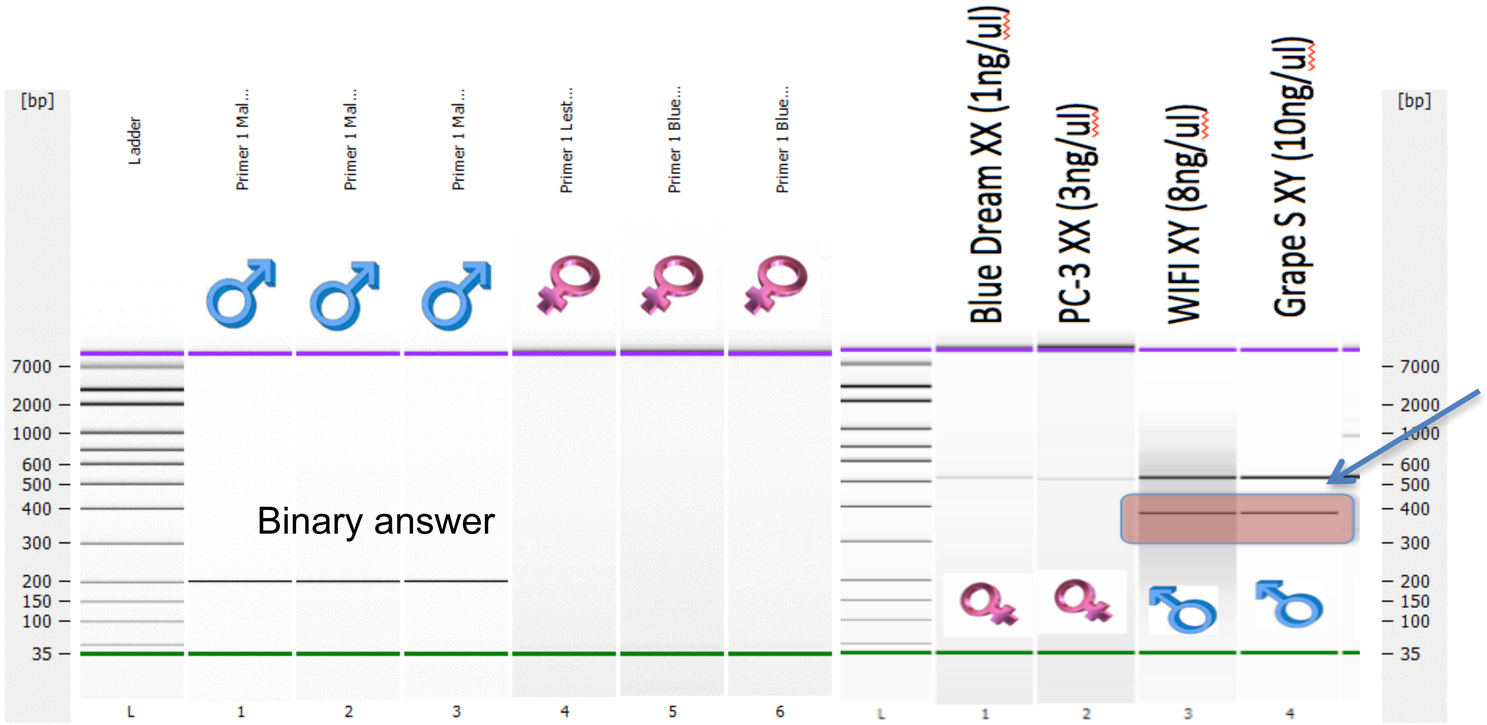
MEDICINAL GENOMICS

Whole genome assembly delivers!

Cleaner Yes/No answer than Mandolino for taqman

Medicinal Genomics- Tadigotla

Mandolino



Males Produce multiple bands

Females produce bands as well

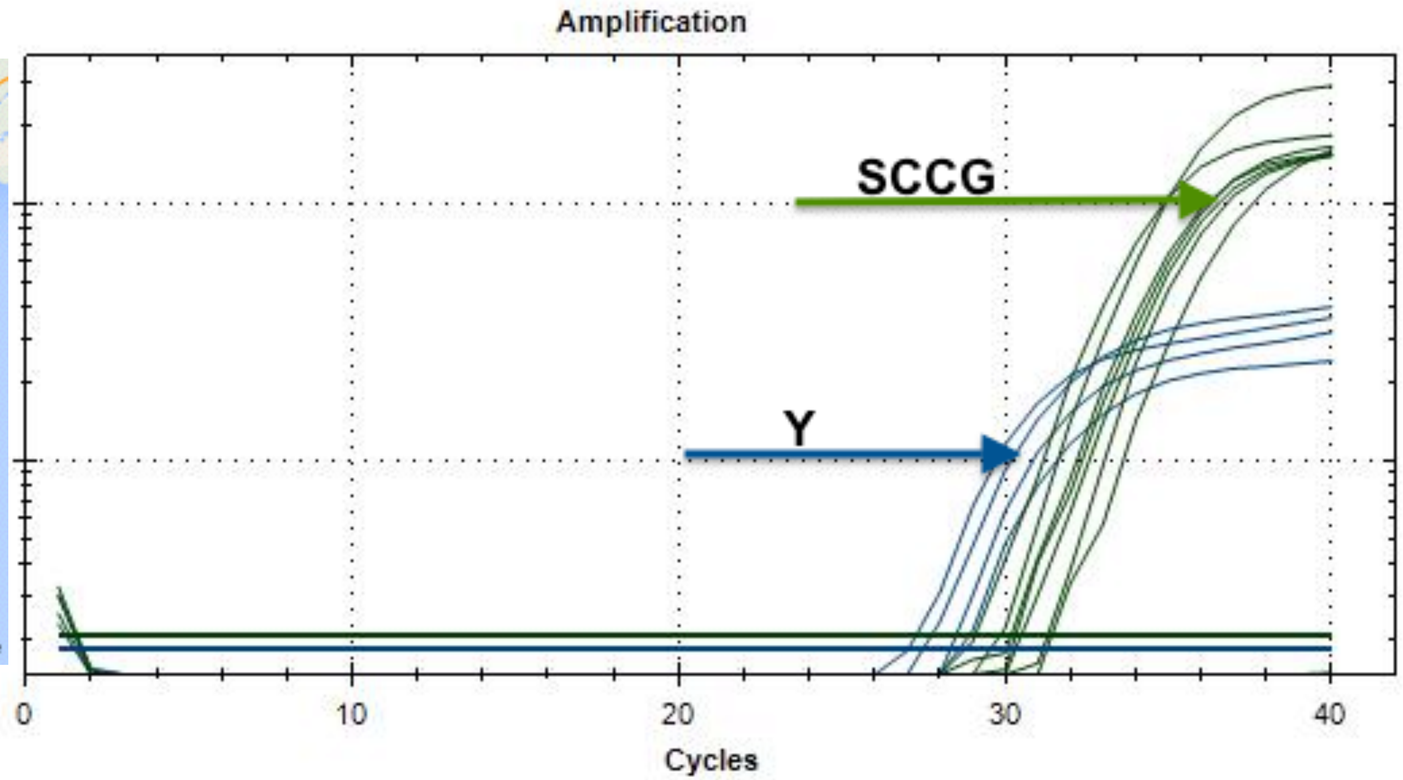
Temperature of PCR changes this!

MEDICINAL GENOMICS

Y amplicon performing as well as diploid SCCG

SCCG is PathogINDICAtor plant control

East Coast Validation (n=10)



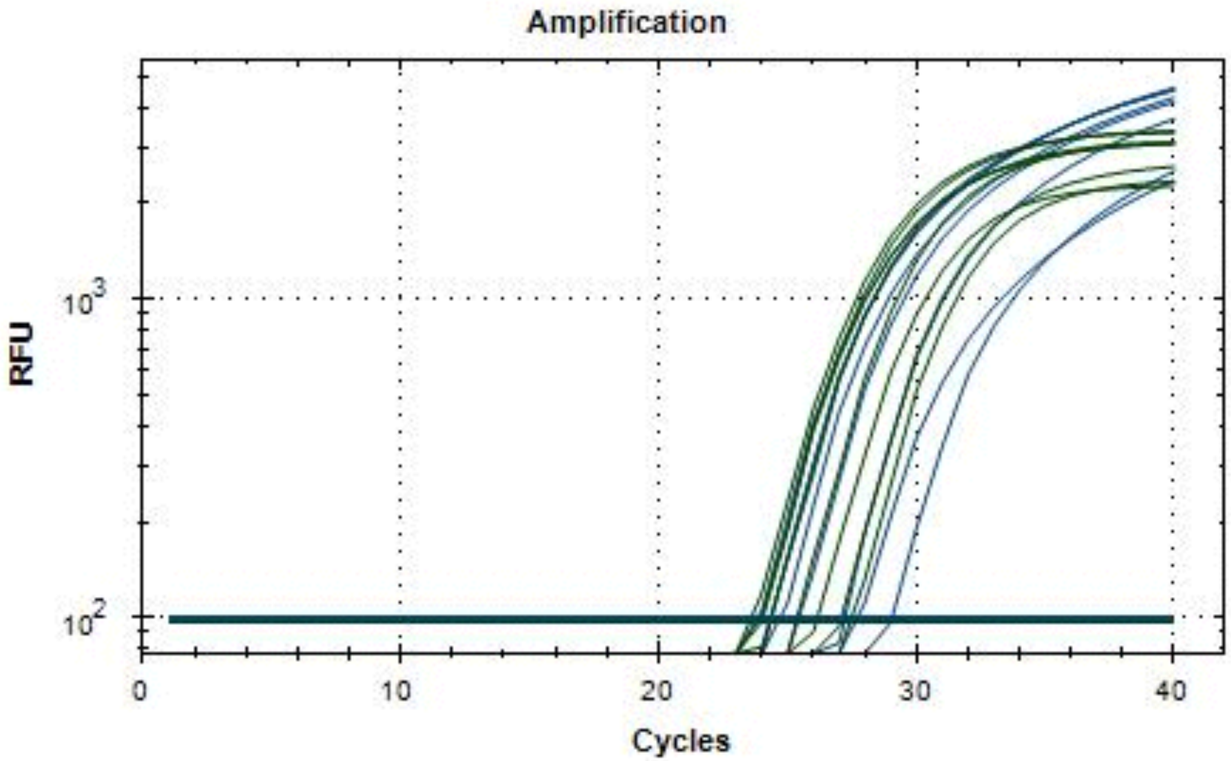
See Tuesday Poster for SCCG/PathogINDICAtor description

MEDICINAL GENOMICS

Replicate with geographically distinct genetics

Pacific Northwest

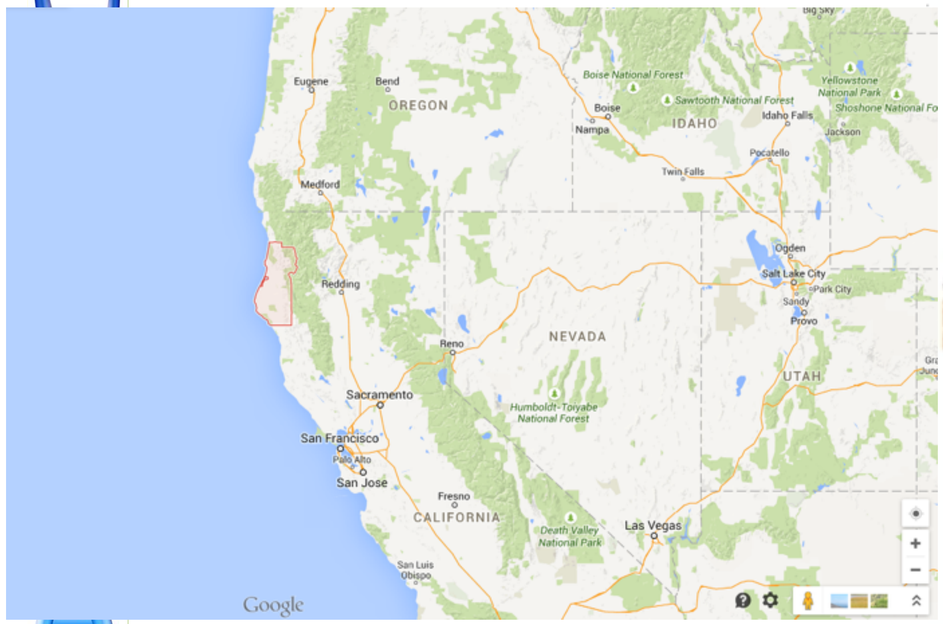
North West Coast Validation (N=10)

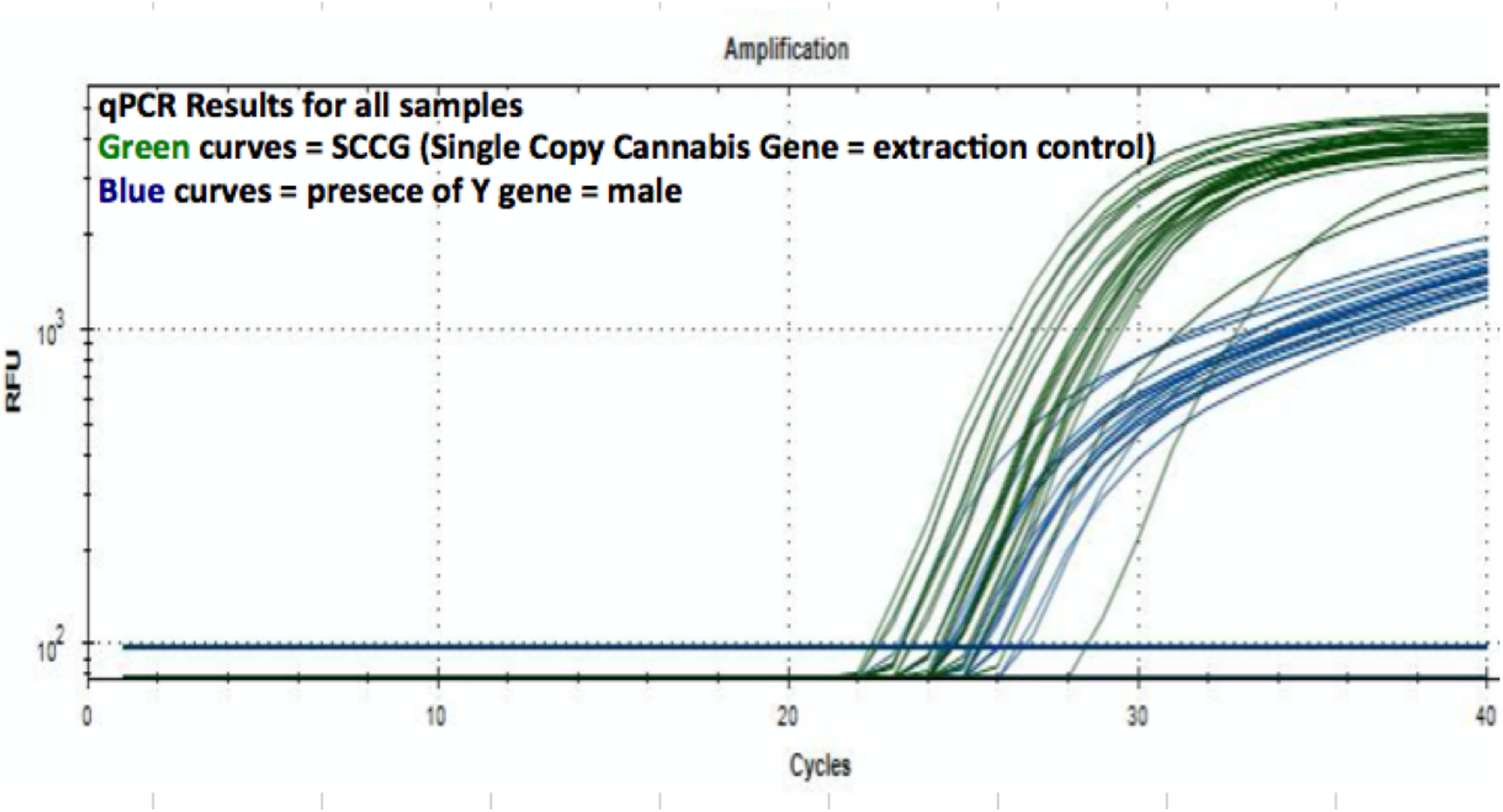


MEDICINAL GENOMICS

Replicate with geographically distinct genetics

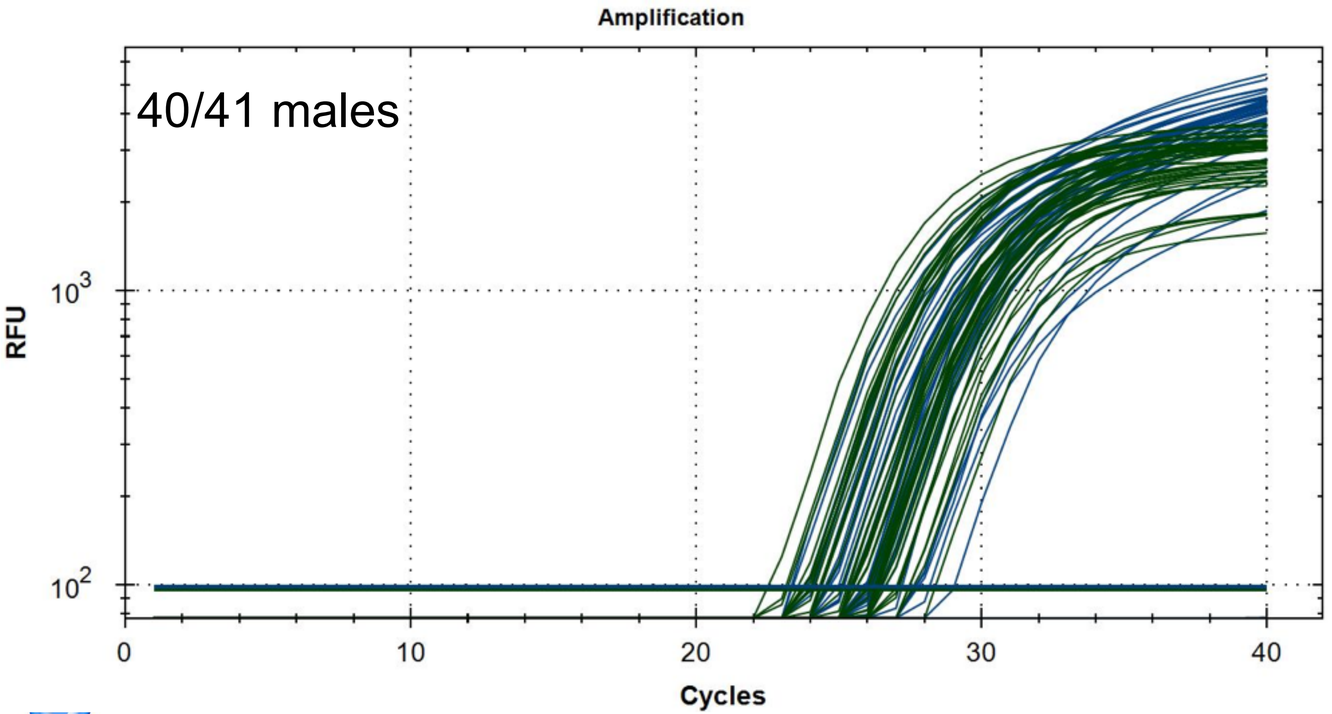
Humboldt County (n=20)



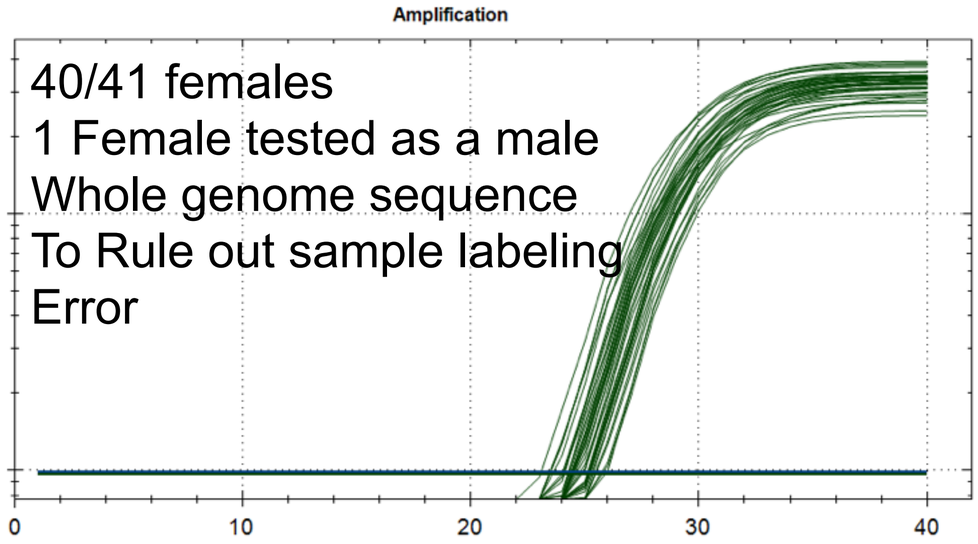


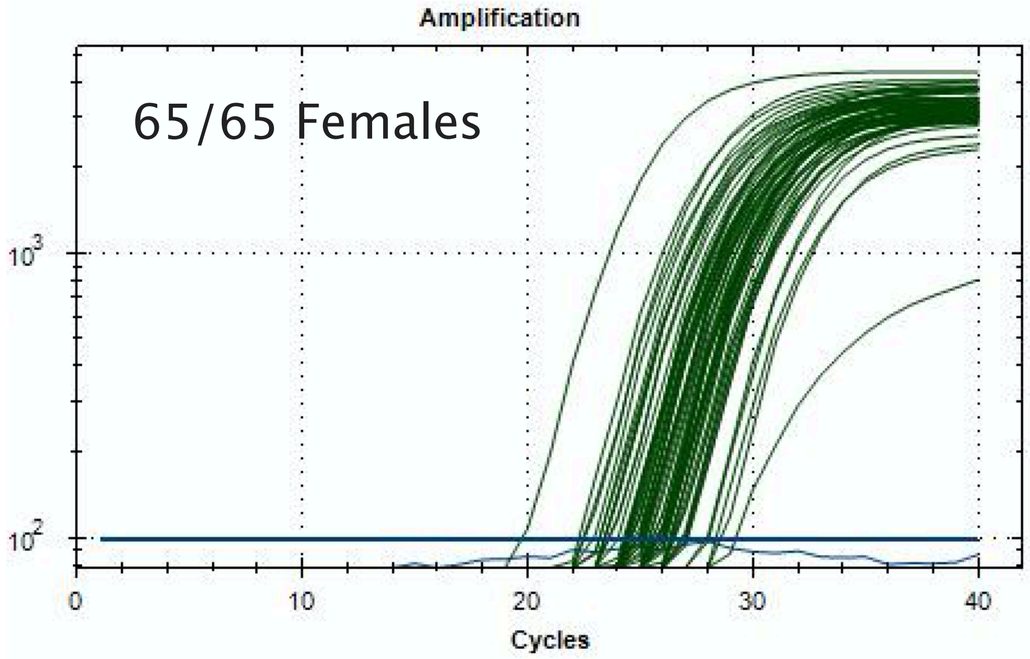
MEDICINAL GENOMICS

All Samples to date



>99% Females test as females



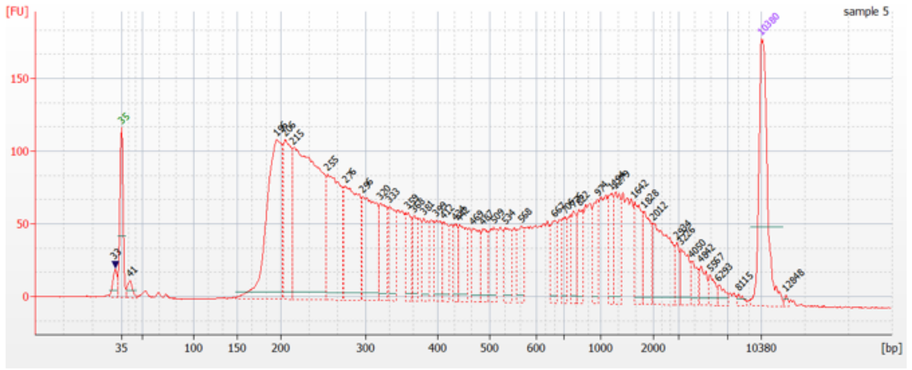


- 1 Male showed no Y signal?
- No Pollen Sack Confirmation from Grower
- Whole Genome Sequence to assess if its an error

- 2 Flowered Hermaphrodites tested as Females

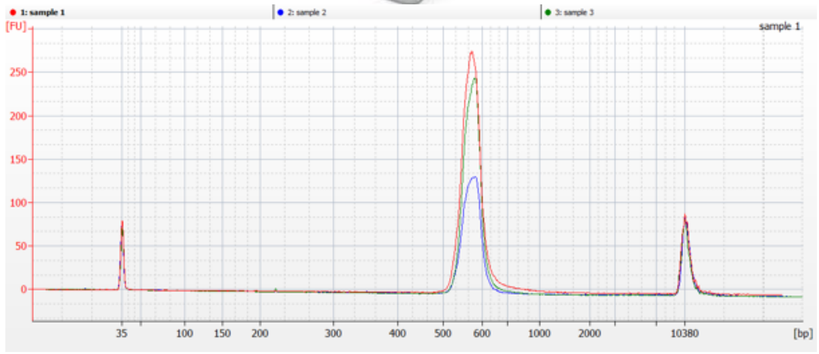
MEDICINAL GENOMICS

Whole Genome Sequencing Discordant Males

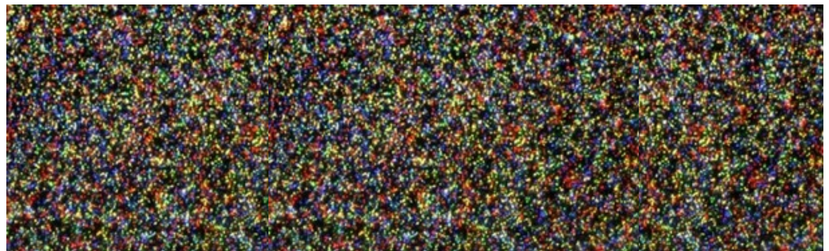


Size Select 500-600bp libraries for 2x250bp sequencing

Measure % Male-Ness
Reads mapping to Putative Male Contigs.



5Gb ILMN Sequencing on Discordant Males

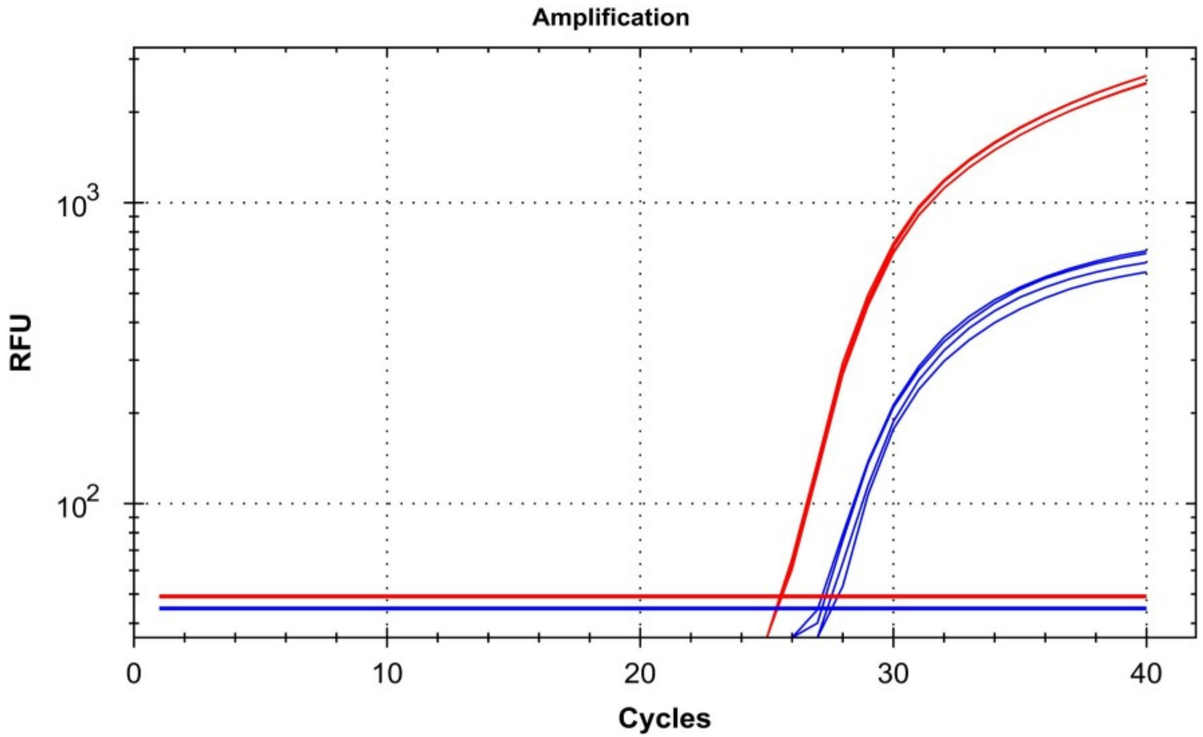


MEDICINAL GENOMICS

Australian Bastard Cannabis (ABCh)

The Most Distant Cultivar we can find (Thank you Werc Shop)

Does it qPCR for Y Chromosome and Cannabinoid gene SCCG in very distant cultivars?



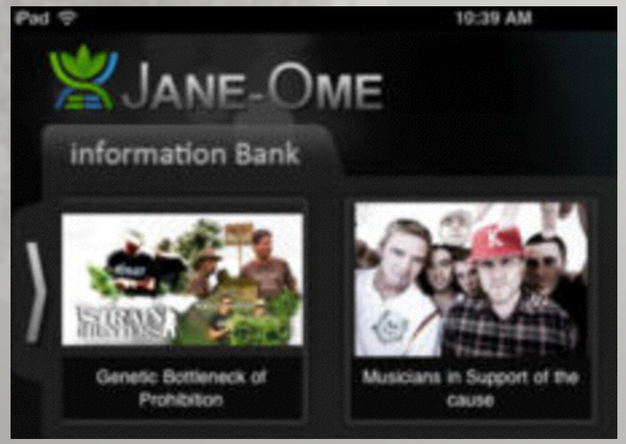
RED= Y Chromosome and Blue = SCCG?

In triplicate

Yes

Support this work

Privately funded
Over 120 investors
Musician in-kind support
Slightly Stoopid



Acknowledgements

Courtagen Team
Yvonne Helbert MS
Lei Zhang PhD
Jessica Spangler MS
Ted Foss PhD
Jason Warner PhD
Vashist Tadigotla PhD
Stephen McLaughlin, PhD
Doug Smith, PhD
Kellie Dodd, MS
Heather Ebling, MS
Matthew Dindinger, MS
John Geanacopoulos

Helpful Communications
Brian Eagar
Steve Amirault
The WercShop
MCR Labs