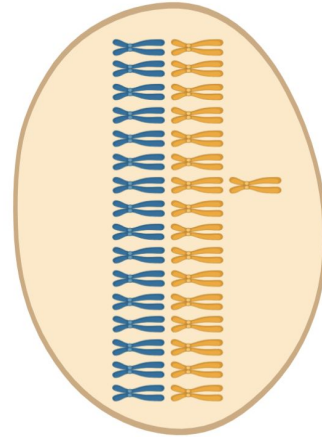


Assessing Ecology and Evolution of Aneuploidy in Yeast

Eduardo Scopel Ferreira da Costa
November 20th 2019



Domesticated

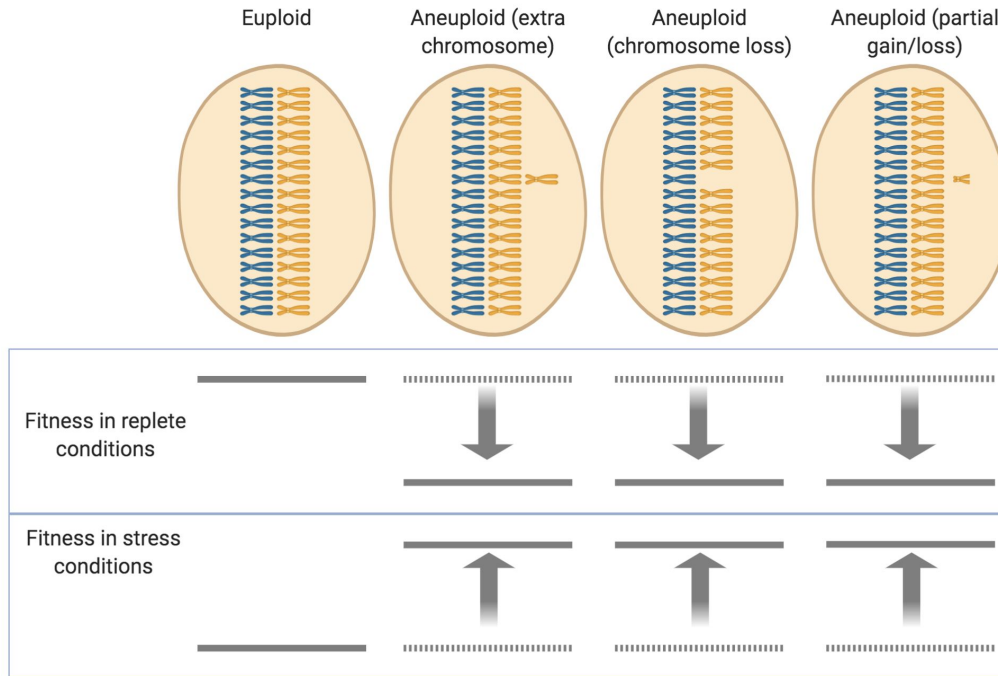
Wild

Human

Unknown

| | | |
|-------|---------------------------------------|-----|
| 01.W2 | Clinical Y ⁺ amplification | 13 |
| 01.W3 | Clinical/S. boulardii | 24 |
| 01.W4 | Georgian | 39 |
| 02.A | Alpechin | 17 |
| 03.B | Brazilian bioethanol | 35 |
| 04.M | Mediterranean oak | 8 |
| 05.F | French dairy | 32 |
| 06.A | African beer | 20 |
| 07.M | Mosaic beer | 21 |
| 08.M | Mixed origin | 72 |
| 09.M | Chinese wine | 16 |
| 10.F | Chinese wine | 13 |
| 11.A | Cocoa wine | 28 |
| 12.C | CHNIII | 2 |
| 13.A | CHNII | 2 |
| 14.C | CHNI | 1 |
| 15.C | Taiwanese | 3 |
| 16.C | Far East Asia (CHNIV) | 9 |
| 17.T | Malaysian | 6 |
| 18.F | CHNV | 2 |
| 19.M | Ecuadorian | 10 |
| 20.C | Far East Russian | 4 |
| 21.E | North America oak | 13 |
| 22.R | Asian island | 11 |
| 23.N | Sake | 47 |
| 24.A | Asian fermentation | 39 |
| 25.S | Mosaic region 1 | 17 |
| 26.A | Mosaic region 2 | 20 |
| M1.M | Mosaic region 3 | 113 |
| M2.M | Unclassified | 48 |
| M3.M | | |
| Unc | | |

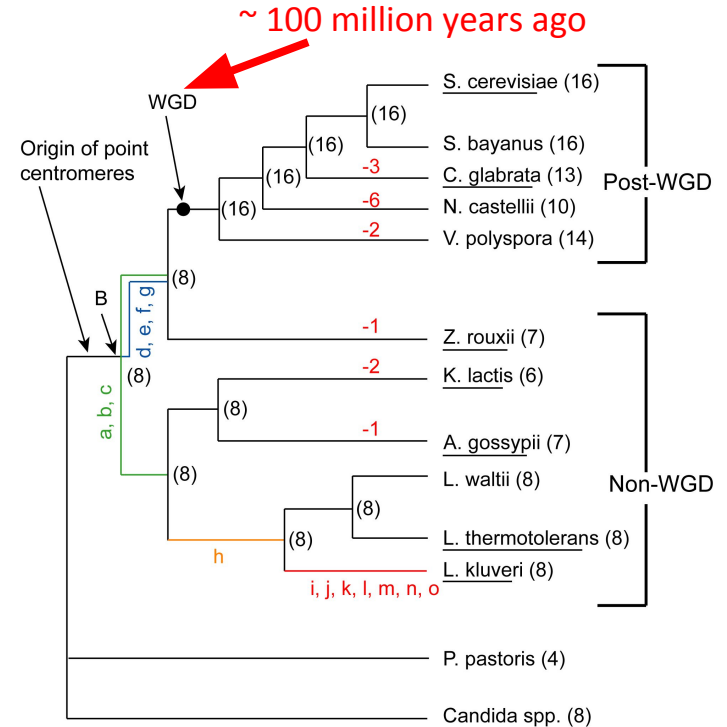
Aneuploidy as a mechanism of local adaptation



- Aneuploidy is usually detrimental to fitness;
- It can increase fitness under specific conditions;
- May be an extreme rapid response to stressful conditions.

Why study aneuploidy in *S. cerevisiae*?

- They live in diverse environments ^[1];
- Many genomes available from around the globe^[1];
- Ploidy change is common, yet the number of chromosomes (16) is conserved^[2,3];
- There is a debate whether aneuploidy is rare^[1] or common^[4] in wild strains.



[1] Peter et al. Nature 556.7701 (2018): 339.

[2] Gordon et al. PLoS genetics 7.7 (2011): e1002190.

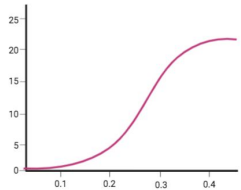
[3] Wolfe & Shields. Nature 387.6634 (1997): 708.

[4] Duan et al. Nature communications 9.1 (2018): 2690.

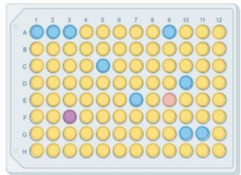
Assessing ecology and evolution of aneuploidy



- **Aim 1:** Evaluate computational tools to detect aneuploidy from yeast genome data.



- **Aim 2:** Identify environmental or genetic factors associated with aneuploidy in *S. cerevisiae*.



- **Aim 3:** Test fitness effects of aneuploidy in different environments.

Aim 1. Using genome data to detect aneuploidy



- Methods used to detect aneuploidy:



Karyotyping;



Flow Cytometry;



qPCR;



Pulsed Field Gel Electrophoresis (PFGE);



Read Depth;



Allele Ratio - also known as B-Allele Frequency (BAF).

- Most of the past work focused on cancer, embryos, CNVs, exome data (known overall ploidy).

Aim 1. Assessing 6 tools with 46 yeast strains



| Tool | Method | Citations | Original Application | Reference |
|----------------|---------|-----------|----------------------|-----------|
| VarScan2 | RD | 2,310 | Cancer | [1] |
| Control-Freec | RD, BAF | 386 | Cancer | [2] |
| nQuire | BAF | 14 | Fungi | [3] |
| ploidyNGS | BAF | 16 | Fungi | [4] |
| Vcf2alleleplot | BAF | 14 | Yeast | [5] |
| Ymap | RD, BAF | 30 | Yeast | [6] |

23 *C. albicans* and 23 *S. cerevisiae* strains:

- 12 aneuploids and 11 euploids for each;
- Previously sequenced;
- Known ploidy (FC);
- Varied genome structure.

[1] Koboldt et al. Genome research 22.3 (2012): 568-576.

[2] Boeva et al. Bioinformatics 28.3 (2011): 423-425.

[3] Weiß et al. BMC bioinformatics 19.1 (2018): 122.

[4] dos Santos et al. Bioinformatics 33.16 (2017): 2575-2576.

[5] Bensasson et al. Genetics 211.1 (2019): 277-288.

[6] Abbey et al. Genome medicine 6.11 (2014): 100.

Aim 1. Accuracy of RD and BAF differs between species



| Tools | VarScan2 RD | Control-Freec RD | nQuire ^{BAF} | ploidyNGS ^{BAF} | vcf2alleleplot ^{BAF} | Ymap ^{RD,BAF} |
|--|------------------------|-----------------------------|-----------------------|--------------------------|-------------------------------|------------------------|
| <i>Saccharomyces cerevisiae</i> (N = 23 genomes, 368 chromosomes, 25 aneuploidies [6.8%]) | | | | | | |
| Sensitivity (TP) | 84% (21) | 94% (23.5) | 76% (17) | 60% (15) | 64% (16) | 82% (20.5) |
| Specificity (TN) | 100% (343) | 98% (336) | 69% (239.5) | 99% (341) | 100% (343) | 99% (340) |
| FDR (FP) | 0% (0) | 23% (7) | 86% (106) | 12% (2) | 0% (0) | 13% (3) |
| <i>Candida albicans</i> (N = 23 genomes, 184 chromosomes, 14 aneuploidies [7.6%]) | | | | | | |
| Sensitivity (TP) | 86% (12) | 86% (12) | 82% (11.5) | 92% (12) | 86% (12) | 86% (12) |
| Specificity (TN) | 100% (170) | 98% (166) | 92% (156) | 99% (169) | 100% (170) | 98% (166) |
| FDR (FP) | 0% (0) | 25% (4) | 55% (14) | 14% (2) | 0% (0) | 20% (3) |

RD - Read Depth tools

^{BAF} - Allele Ratio tools

Sensitivity (*True Positive Rate*) = (True Positives)/(True Positives+False Negatives)

Specificity (*True Negative Rate*) = (True Negatives)/(True Negatives+False Positives)

FDR (*False Discovery Rate*) = (False Positives)/(True Positives + False Positives)

Aim 1. Conclusion



Read Depth tools

- + Not affected by heterozygosity
- + Can detect segmental copy number variation
- Require good sequencing/DNA quality
- Do not detect overall ploidy

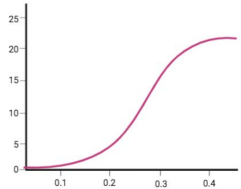
Allele Ratio

- + Detect overall ploidy
- + Only few HQ heterozygous SNPs are enough
- + Work well on low quality DNA
- Only work on heterozygous strains
- Not good for CNVs

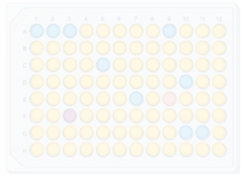
Assessing ecology and evolution of aneuploidy



- **Aim 1:** Evaluate computational tools to detect aneuploidy from yeast genome data.

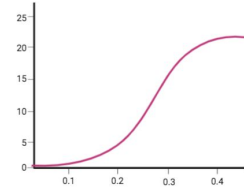


- **Aim 2:** Identify environmental or genetic factors associated with aneuploidy in *S. cerevisiae*.



- **Aim 3:** Test fitness effects of aneuploidy in different environments.

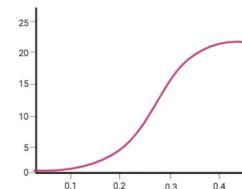
Aim 2. Factors associated with aneuploidy



What do environments with high frequency of aneuploidy have in common?

Are there genetic factors associated with high frequency of aneuploidy?

Aim 2. Sorting strains into 19 environments



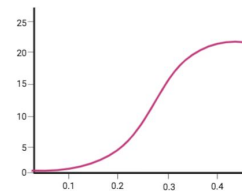
| Environment | # of strains |
|---|--------------|
| Wine | 363 |
| Commercial (including starter cultures) | 223 |
| Beer | 184 |
| Liquid Fermentation | 136 |
| Bread | 129 |
| Sake | 59 |
| Dairy | 47 |
| Solid-state fermentation (SSF) | 45 |
| Bioethanol | 43 |
| Lab | 28 |
| Coffee | 22 |
| Cocoa | 18 |
| Industrial | 10 |
| Clinical | 277 |
| Trees | 192 |
| Fruit | 156 |
| Flower | 74 |
| Insect | 26 |
| Other plants | 16 |
| Total | 2048 |

Human-associated (n=1307)

Clinical (n=277)

Wild (n=464)

Aim 2. Environmental differences in aneuploidy



- Aneuploidy is **common** regardless of environment;
- Aneuploidy **differs** among environments;
- **Domestication** does **not** explain the differences.

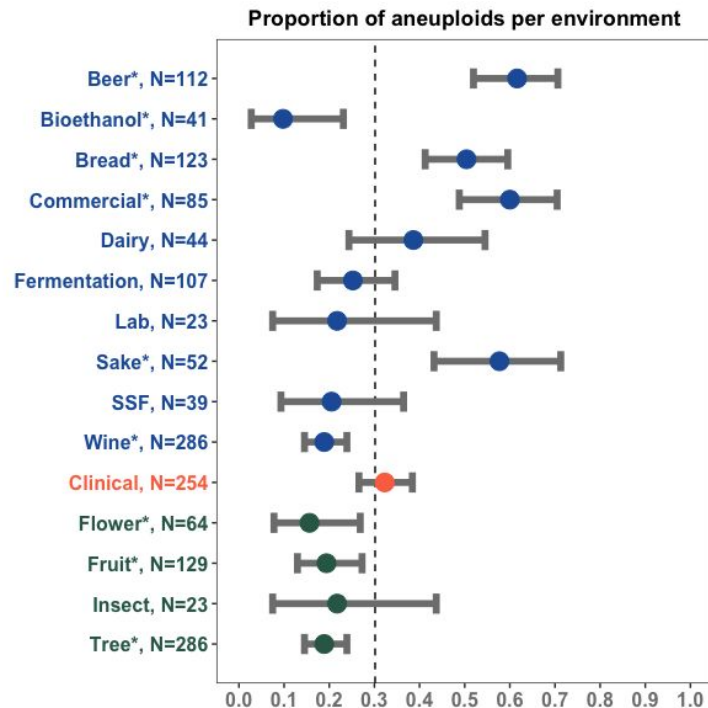
I designed a logistic model (**glm**) to test for differences in frequency of aneuploidy:

Aneuploidy ~ Environment * heterozygosity

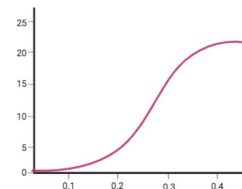
↓
Binary
[Yes..No]

↓
Categorical
[14 factors]

↓
Continuous



Aim 2. aneuploidy ~ environment + heterozygosity



After the following simplifications:

1. Grouping **Beer** and **Sake** into **Brewing**;
2. Grouping **Flower**, **Fruit**, **Insect**, **Tree**, and **Wine** into **Seasonal (or diurnal)**;
3. Removing the interaction term;

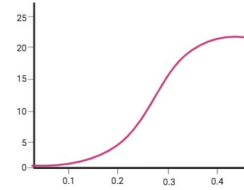
The final model explains **28%** (glm, n=773, df=10, $p < 2.2e-16$) of the deviance (environment, 21%; heterozygosity, 7%):

aneuploidy ~ environment + heterozygosity

I have tried other simplifications that resulted in worse models:

1. Grouping **Wild**, **Domesticated** and **Clinical** strains (3 environments);
2. Grouping Open Fermentation strains: **Bioethanol**, **Dairy**, **Wine** and **Wild** (7 environments);
3. Grouping Non-Seasonal strains: **Brewing** and **Commercial** (7 environments);
4. Grouping Continuous Growth strains: **Bioethanol** and **Brewing** (8 environments);
5. Removing heterozygosity or environment.

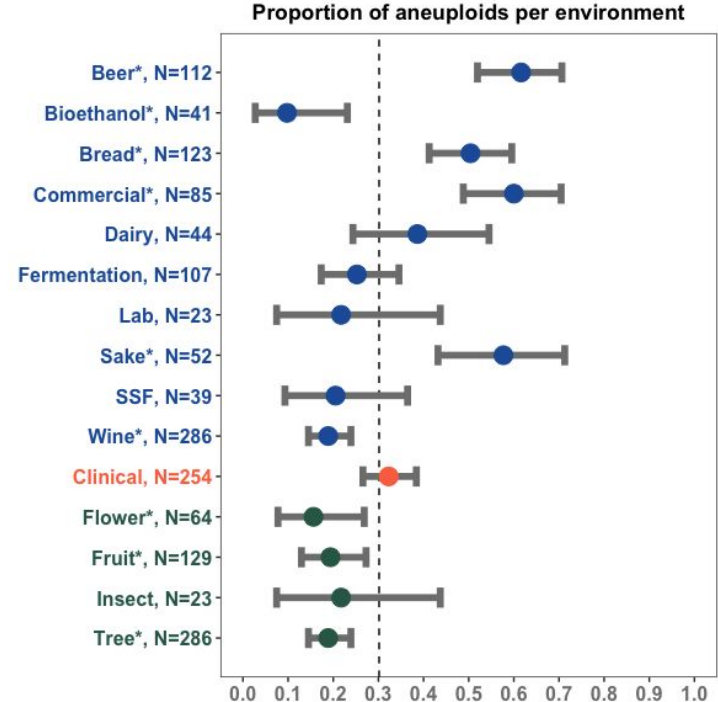
Aim 2. Environmental differences in aneuploidy



What happens in **seasonal (or diurnal)** environments?



Low aneuploidy in **domesticated wine** and **wild** environments suggests **meiosis is a better predictor of aneuploidy** than domestication.



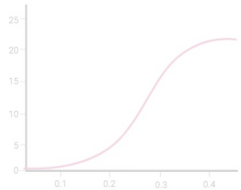
Aim 2. Next steps

- Include lineage (genetic background) as an explanatory variable;
- Consider removing highly heterozygous strains;
- Identify what chromosomes are more frequently gained or lost;
- Annotate a phylogenetic tree with each aneuploidy event found here.

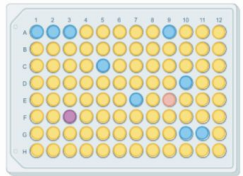
Assessing ecology and evolution of aneuploidy



- **Aim 1:** Evaluate computational tools to detect aneuploidy from yeast genome data.

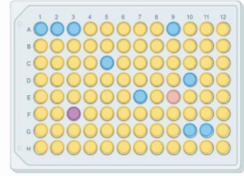


- **Aim 2:** Identify environmental or genetic factors associated with aneuploidy in *S. cerevisiae*.



- **Aim 3:** Test fitness effects of aneuploidy in different environments.

Aim 3. Fitness effects of aneuploidy



What are the **fitness effects** (growth, sporulation) of **aneuploidy** in different lineages growing under **distinct environments**?

Aim 3a. Phenotypic differences: ploidy and lineage

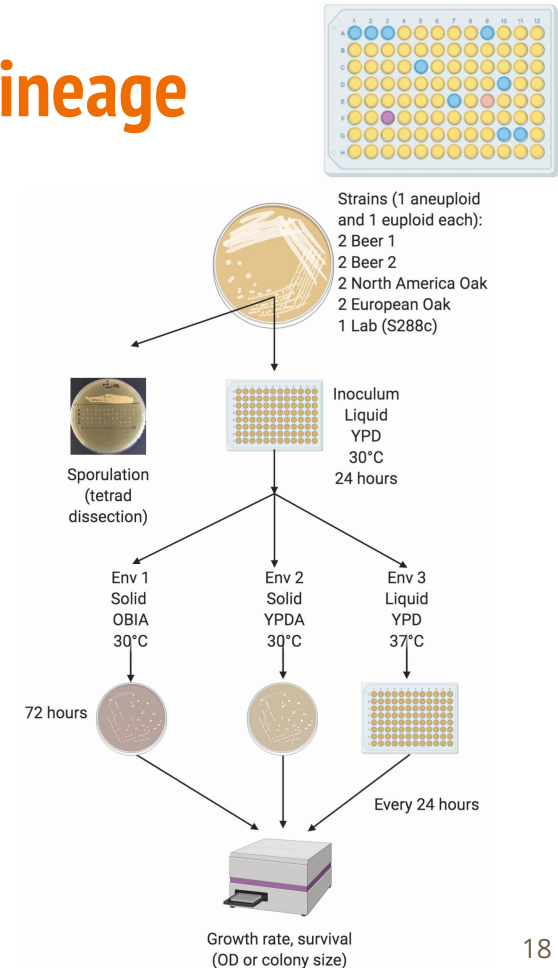
Measure **phenotypic** differences between **aneuploid** and **euploid** strains living in **wild** and **beer**-like environments.

Using **human** and **non-human** associated lineages:

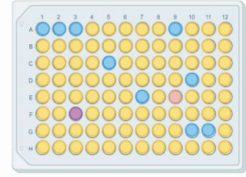
- **North America Oak, European Oak;**
- **Lab (S288c);**
- **Beer 1, Beer 2.**

In 3 environments:

- **Tree-like (Oak Bark Infusion Agar);**
- **Solid control (YPD Agar);**
- **Beer-like (Liquid YPD).**



Aim 3b. Evolving wild yeast under distinct conditions



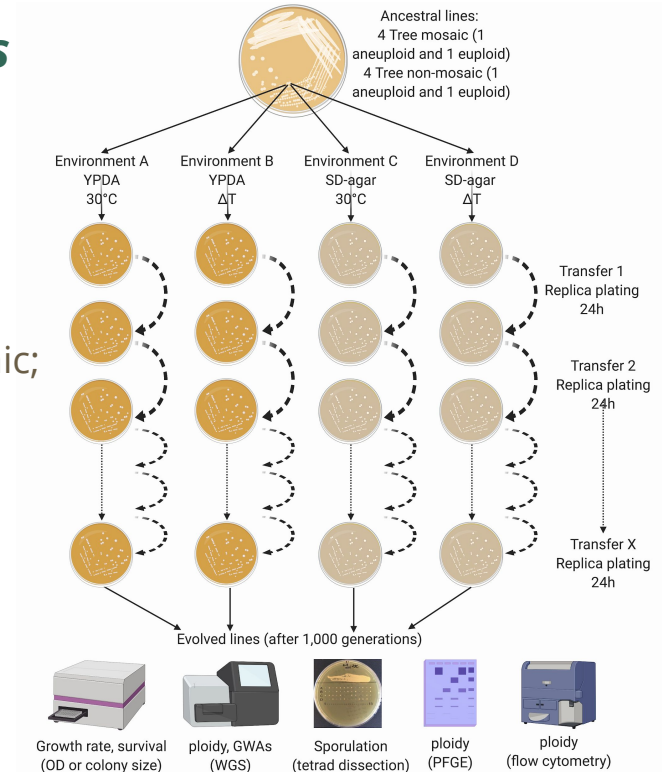
Determine how **aneuploid** and **euploid wild strains** respond to **wild** and **beer-like** environments.

Using **Wild** and **Lab** lineages:

- 2 North America Oak and 2 European Oak mosaic;
- 2 North America Oak and 2 European Oak non-mosaic;
- 1 Lab (S288c).

Evolve for 1,000 generations in 4 environments:

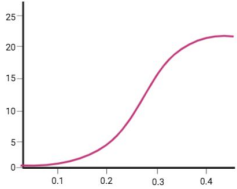
- Beer-like** (nutrient-rich, stable temperature);
- Fluctuating temperature**;
- Nutrient-poor**;
- Tree-like** (fluctuating temperature & nutrient-poor).



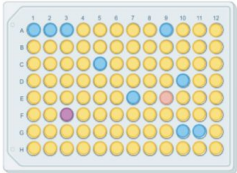
Summary



Aim 1. Assess computational tools to detect aneuploidy using yeast genome data;



Aim 2. Explain the association between frequency of aneuploidy and environmental or genetic factors in yeast



Aim 3a. Test fitness consequences of aneuploidy in different yeast lineages under distinct environments;

Aim 3b. Determine how aneuploid and euploid wild yeast strains evolve under different conditions.

Association between aneuploidy and environment?

- Aneuploidy is common in clinical^[1], sake, beer, and admixed^[2] strains of *S. cerevisiae*;
- However, there is a debate whether the frequency of aneuploidy in wild strains, is rare^[2] or common^[3].

[1] Zhu et al. G3: Genes, Genomes, Genetics (2016): g3-116.

[2] Peter et al. Nature 556.7701 (2018): 339.

[3] Duan et al. Nature communications 9.1 (2018): 2690.

Ploidy and heterozygosity affect aneuploidy

