Patterns of gene duplication and sex chromosomes evolution

Esther Betrán
University of Texas-Arlington
Outline:

Introduction:
- Sex chromosomes evolution
- Retroposition as a gene duplication mechanism

Recent surprises:
- Rates and patterns of gene duplication change the traditional view of X chromosome evolution
- Three hypotheses to explain the patterns
  - X inactivation
  - Sexual antagonism
  - Meiotic drive
- Some surprises also in the Y chromosome.

Summary
Human sex chromosomes
Human sex chromosomes

Mostly heterochromatin
Human sex chromosomes X and Y are morphologically very different: Y is small ~ 78 genes and X is big ~1200 genes.
However they begun their journey as a normal pair of autosomes where an allele of a gene (SRY) evolved to have the sex determining role.
SoX3

Autosomes
Genic sex determination

Autosomes

Proto X  Proto Y

SoX3  SRY
Autosomes → Proto X Proto Y → X

Chromosomal sex determination
Proto X chromosome

Proto Y chromosome

SoX3

SRY

Gene expressed in both sexes
Proto X chromosome

Proto Y chromosome

SoX3

SRY

- **Blue** Allele that benefits the males but hurts the females
- **Pink** Allele that benefits the females but hurts the males
Proto X chromosome

SoX3

Proto Y chromosome

An inversion would be beneficial

Allele that benefits the males but hurts the females

Allele that benefits the females but hurts the males
Proto X chromosome

Allele that benefits the males but hurts the females

Proto Y chromosome

SoX3

SRY

Recombination stops and the region of the Y degenerates

- Blue: Allele that benefits the males but hurts the females
- Pink: Allele that benefits the females but hurts the males
Proto X chromosome

Allele that benefits the males but hurts the females

Proto Y chromosome

SRY

SoX3

In addition the genes in that region begin evolving separately

Rice 1996 for a review

- **Blue**: Allele that benefits the males but hurts the females
- **Pink**: Allele that benefits the females but hurts the males
Degeneration of the Y chromosome

Several mechanisms related to the lack of recombination:

1. One process is deleterious mutation accumulation by Muller's ratchet, leading to an increasing number of deleterious mutations, which become fixed as the process continues and they cannot be recombined out.

2. Another possibility is hitch-hiking: favorable mutant alleles arise on the proto-Y and rise in frequency to fixation, concomitantly fixing deleterious alleles on the same chromosome.

3. Background selection, selection against strongly deleterious mutations, will have the effect of reducing the population size. This accelerates the fixation of mildly deleterious mutations and reduces the chance of fixation of mildly advantageous mutations.

Charlesworth and Charlesworth 2000 for a review
In summary, the nonrecombining Y chromosome will accumulate deleterious mutations. A fraction of those will be caused by insertions of transposable elements (TE).
Four strata in the human Y chromosome

Lahn and Page 1999
At the end the non-recombining region of the Y chromosome is left with very few genes that are male-specific.

Y chromosome specialization
Dosage compensation is the mechanism by which the genes on the X chromosome express at the same level in male and female.

Females are XX

Males are XY
Traditional view:

- Y chromosome degenerates and specializes
- X chromosome undergoes dosage compensation
- X chromosome conserved gene content
- favored location for male- and female-biased genes
Gene duplication:

a. genome duplication
b. segmental duplication
   - tandem duplication
   - transposition
c. retroposition

Mechanisms reviewed in:
Long et al. Nature Reviews Genetics 2003
New gene formation by retroposition
Consequences of retroposition

- Hallmarks of retroposition:
  - intronless
  - poly-A tract
  - flanking direct repeats
- Parental and derived genes have different location
- Parental and derived genes share sequence similarity
- The retrogene usually lacks regulatory region at the time of insertion
Direction of copying is clear

Parental gene >1 exon

Retrogene 1 exon

e.g. X 3
We have studied retrogenes and patterns of duplication in:

**Fly**
(Betrán et al. Genome Research 2002; Bai et al. Genome Biology 2007)

**Human and Mouse**
(Emerson et al. Science 2004)

**Chicken**

**Worm**
(In preparation)
Drosophila

We have now 12 sequenced and annotated genomes of fruitflies and is serving as a good model system.
Identifying retrogenes in the *Drosophila* genome
D. melanogaster chromosomes
D. melanogaster chromosomes

Y
X
2
3
4
2L
2R
3L
3R
D. melanogaster chromosomes

94 retroposition events
Rate of retroposition in the *Drosophila* genome
Rate of retroposition in the *Drosophila* genome
Rate of duplication

- 32 retrogenes originated in the last 63 My

(Tree from Tamura et al. 2004)

Bai et al. Genome Biology 2007
Rate of duplication

- 32 retrogenes originated in the last 63 My
- In the *D. melanogaster* lineage, the rate is 1 gene every two My.
- The rate seems to be constant and ongoing
- Similar to human lineage; 1 retrogene per My (Marques et al. 2005)
- Small fraction (3%) of all duplications in Drosophila (17 genes per My; Hahn et al. 2007)

(Tree from Tamura et al. 2004)

Bai et al. Genome Biology 2007
### Table 1. Analysis of duplication between chromosomes. Expected values were calculated following Betrán et al. (2002)

<table>
<thead>
<tr>
<th>Direction</th>
<th>Expectation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>No</td>
</tr>
<tr>
<td>X → A</td>
<td>23.3</td>
<td>13.5</td>
</tr>
<tr>
<td>A → X</td>
<td>20.3</td>
<td>11.8</td>
</tr>
<tr>
<td>A → A</td>
<td>56.4</td>
<td>32.7</td>
</tr>
</tbody>
</table>

\[X^2 = 27.0496; \text{df} = 2; P = 0.000001\]

X, X chromosome; A, Autosome.
### TE distribution in the *D. melanogaster* euchromatin

<table>
<thead>
<tr>
<th>Row</th>
<th>Factors</th>
<th>df</th>
<th>Coefficient Sign</th>
<th>$F$</th>
<th>$p(F)$</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Recombination</td>
<td>1</td>
<td>—</td>
<td>1,283.4</td>
<td>&lt;0.001</td>
<td>17.4%</td>
</tr>
<tr>
<td>b</td>
<td>Intergenic or (introns + UTRs) length</td>
<td>1</td>
<td>+</td>
<td>1,156.6</td>
<td>&lt;0.001</td>
<td>15.7%</td>
</tr>
<tr>
<td>c</td>
<td>Proportion of conserved sequences</td>
<td>1</td>
<td>—</td>
<td>140.1</td>
<td>&lt;0.001</td>
<td>1.9%</td>
</tr>
<tr>
<td>d</td>
<td>X versus autosomes</td>
<td>1</td>
<td>—</td>
<td>22.2</td>
<td>&lt;0.001</td>
<td>0.3%</td>
</tr>
<tr>
<td>e</td>
<td>Neighborhood</td>
<td>1</td>
<td>+</td>
<td>23.4</td>
<td>&lt;0.001</td>
<td>0.3%</td>
</tr>
<tr>
<td>f</td>
<td>Germline- versus soma-expressed genes</td>
<td>1</td>
<td>—</td>
<td>0.9</td>
<td>0.92</td>
<td>0.0%</td>
</tr>
<tr>
<td>g</td>
<td>Intergenic regions versus genes</td>
<td>1</td>
<td>(−)</td>
<td>37.8</td>
<td>&lt;0.001</td>
<td>0.5%</td>
</tr>
<tr>
<td>h</td>
<td>Neighborhood : germline versus soma</td>
<td>1</td>
<td>(+)</td>
<td>2.7</td>
<td>0.1</td>
<td>0.0%</td>
</tr>
<tr>
<td>i</td>
<td>Neighborhood : intergenic regions versus genes</td>
<td>1</td>
<td>—</td>
<td>8.5</td>
<td>&lt;0.01</td>
<td>0.1%</td>
</tr>
<tr>
<td>j</td>
<td>Germline versus soma : intergenic regions versus genes</td>
<td>1</td>
<td>(−)</td>
<td>16.3</td>
<td>&lt;0.001</td>
<td>0.2%</td>
</tr>
<tr>
<td>k</td>
<td>Neighborhood : germline versus soma: intergenic regions versus genes</td>
<td>1</td>
<td>+</td>
<td>5.0</td>
<td>&lt;0.05</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

The variable “neighborhood” is the proportion of germline-expressed genes among the ten neighbors of a focal gene or intergenic region. Coefficient signs of the GLM indicate the sign of linear coefficients associated with the terms of the model and hence indicate the direction (positive or negative) in which the different factors affect the number of TE insertions. Parentheses around the sign indicate that the linear coefficients were not significantly different from zero in the model.

doi:10.1371/journal.pgen.0030210.t001

Fontanillas *et al.* 2007
### TE distribution in the *D. melanogaster* euchromatin

<table>
<thead>
<tr>
<th>Row</th>
<th>Factors</th>
<th>df</th>
<th>Coefficient Sign</th>
<th>$F$</th>
<th>$p(F)$</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Recombination</td>
<td>1</td>
<td>$-$</td>
<td>1,283.4</td>
<td>$&lt;0.001$</td>
<td>17.4%</td>
</tr>
<tr>
<td>b</td>
<td>Intergenic or (introns + UTRs) length</td>
<td>1</td>
<td>$+$</td>
<td>1,156.6</td>
<td>$&lt;0.001$</td>
<td>15.7%</td>
</tr>
<tr>
<td>c</td>
<td>Proportion of conserved sequences</td>
<td>1</td>
<td>$-$</td>
<td>140.4</td>
<td>$&lt;0.001$</td>
<td>1.0%</td>
</tr>
<tr>
<td>d</td>
<td>X versus autosomes</td>
<td>1</td>
<td>$-$</td>
<td>22.2</td>
<td>$&lt;0.001$</td>
<td>0.3%</td>
</tr>
<tr>
<td>e</td>
<td>Neighborhood</td>
<td>1</td>
<td>$+$</td>
<td>25.4</td>
<td>$&lt;0.001$</td>
<td>0.3%</td>
</tr>
<tr>
<td>f</td>
<td>Germline- versus soma-expressed genes</td>
<td>1</td>
<td>$-$</td>
<td>0.9</td>
<td>0.92</td>
<td>0.0%</td>
</tr>
<tr>
<td>g</td>
<td>Intergenic regions versus genes</td>
<td>1</td>
<td>$-$</td>
<td>37.8</td>
<td>$&lt;0.001$</td>
<td>0.5%</td>
</tr>
<tr>
<td>h</td>
<td>Neighborhood : germline versus soma</td>
<td>1</td>
<td>$+$</td>
<td>2.7</td>
<td>0.1</td>
<td>0.0%</td>
</tr>
<tr>
<td>i</td>
<td>Neighborhood : intergenic regions versus genes</td>
<td>1</td>
<td>$-$</td>
<td>8.5</td>
<td>$&lt;0.01$</td>
<td>0.1%</td>
</tr>
<tr>
<td>j</td>
<td>Germline versus soma : intergenic regions versus genes</td>
<td>1</td>
<td>$-$</td>
<td>16.3</td>
<td>$&lt;0.001$</td>
<td>0.2%</td>
</tr>
<tr>
<td>k</td>
<td>Neighborhood : germline versus soma: intergenic regions versus genes</td>
<td>1</td>
<td>$+$</td>
<td>5.0</td>
<td>$&lt;0.05$</td>
<td>0.1%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>15,674</td>
<td></td>
<td></td>
<td></td>
<td>36.5%</td>
</tr>
</tbody>
</table>

The variable “neighborhood” is the proportion of germline-expressed genes among the ten neighbors of a focal gene or intergenic region. Coefficient signs of the GLM indicate the sign of linear coefficients associated with the terms of the model and hence indicate the direction (positive or negative) in which the different factors affect the number of TE insertions. Parentheses around the sign indicate that the linear coefficients were not significantly different from zero in the model.

doi:10.1371/journal.pgen.0030210.t001

*Fontanillas et al. 2007*
Retrogenes express highly in testis

- 51% of retrogenes express predominantly in testis
- 36% are uniquely expressed in testis
- Genome percentage of uniquely expressed in testis is 7%

Bai et al. Genome Biology 2007
Bai et al. BMC Genomics 2008
Humans & Mouse

- Human and mouse are ideal because they have a lot of retrogene and retropseudogenes
- There is a bias for male germline expression
- Ongoing duplication pattern

95% prediction interval

Emerson et al. Science 2004
Export started with sex chromosomes birth

Potrzebowski et al 2008
Data changes previous believes

- In every genome where we have observed enough “genes going retro” we see an ongoing X chromosome export of male-biased genes.

- This observation is against the traditional believe that male-biased genes should be on the X chromosome.

More dynamic view of X chromosome evolution.
Why excess export from the X?

Hypothesis I: Avoidance of male meiotic X-inactivation

– Generating a copy elsewhere allows genes to avoid silencing during spermatogenesis

– Many of the genes that produce copies are essential genes and this avoidance could provide a selective advantage

X inactivation reduces transcription in testis

Believed to take place to protect unpaired sex chromosomes and/or suppress recombination between sex chromosomes

Richler et al. 1992 Nature Genetics
Predictions from X inactivation

- Intense selection for essential genes. A lot of pressure early on and less pressure now.

- It is a type of subfunctionalization (i.e. partition of the original pattern of expression)

- Genes should keep original function (i.e. be under quite strong purifying selection)

- Genes will transcribe during X inactivation and mutations will cause sterility
Mammalian male germline X postmeiotic reactivation

Single gene studies, and microarray and in situ hybridization of nascent transcripts in postmeiotic cells. Many multicopy genes are expressed from the X.

Wang et al. 2001; Mueller et al. 2008
Hypothesis II: Sexual antagonism

- Female can select for genes in the X chromosome because it spends 2/3 of the time in females
- Male-specific genes avoid X chromosome in general in male germline but also somatic tissues

Supported by the autosomal location of many male germline and male somatic genes: Parisi et al. 2003 and 2004; Sturgill et al. 2007; Zhang et al. 2007.

Reviewed in Betrán et al. Cell Cycle 2004
Genes under sexual antagonism?

Male and female have the same genome (i.e. very few genes are on the Y chromosome) but different morphology and deployment of that genome.

In antagonistic genes, there are allele/s that benefit the female but harm the male and allele/s that are better for the male but not good for the females. Selection acting on these alleles depends on their dominance, sex effect and chromosomal location.
In antagonistic genes, there are allele/s that benefit the female but harm the male and allele/s that are better for the male but not good for the females. Selection acting on these alleles depends on their dominance, sex effect and chromosomal location.
Predictions from sexual antagonism

- Intense selection for antagonistic genes (mostly on the X chromosome)
- It is a type of subfunctionalization but there should be specialization after duplication
- Genes should not exactly keep original function but similar
- X should be a disfavored location in male tissues at all times (all the time during meiosis and somatic cells)
Germline expression

Lack of genes on the X-chromosome

Excess of ovary biased genes on the X

Parisì et al. 2003
As before, lack of genes on the X-chromosome X demasculinization

Parisi et al. 2003
Male-biased genes in Drosophila

• Excess (7-14%) compare to female-biased (3-9%) genes in all lineages
• High birth and extinction rates
• Higher divergence
• Autosomal location and X chromosome demasculinization (30-43% less male-biased genes than expected under uniform distribution). Seen in neo-sex chromosomes in less than 12 My in *D. pseudoobscura* because of new male-biased genes in autosomes.

Let me take a detour!
Recurrence retroposition events from X → A

- D. melanogaster
- D. simulans
- D. sechellia
- D. yakuba
- D. erecta
- D. ananassae
- D. pseudoobscura
- D. persimilis
- D. willistoni
- D. mojavensis
- D. virilis
- D. grimshawi

- Yellow squares: Retroposition events of Ran (CG1404)
- White squares: Retroposition events of Dntf-2 (CG1740)

Bai et al. Genome Biology 2007
Recurrent retroposition events from X → A

Bai et al. Genome Biology 2007
Nuclear transport overview

Figure from Isgro and Schulten, 2007
Rate of evolution

PAML Ka/Ks

Dntf2: 0.0241
Ran: 0.0044

Tracy et al. In preparation
### Ran-like interactions

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ran_H.sapiens</em></td>
<td>MAAGCRPQVQFVILVLCGCKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran_C.familiaris</em></td>
<td>MAAGGEPQVQFVILVLCGCKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran_D.sim</em></td>
<td>MAACQDIPQTCKVILCNGCKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran_D.sech</em></td>
<td>MAACQDIPQTCKVILCNGCKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran_D.ericola</em></td>
<td>MAACQDIPQTCKVILCNGCKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran-like_D_sim</em></td>
<td>MQOEERVKATKIFKSLILCFEEKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran-like_D_sech</em></td>
<td>MQOEERVKATKIFKSLILCFEEKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
<tr>
<td><em>Ran-like_D_mel</em></td>
<td>MQOEERVKATKIFKSLILCFEEKTTTVHENRSTGFREEVEHLLFVHFRHRGPIK 60</td>
</tr>
</tbody>
</table>

**KT Interaction with RCC1**

**KT Switch I**
- Interaction with Dntf-2
- Interaction with RanGap
- Interaction with exportins

**KT Switch II**
- Interaction with Importin β
- C terminal end interactions

Arrows show positively selected sites inferred using PAML

Tracy et al. In preparation
Ran-like interactions

- Interaction with RCC1
- Switch I
- Interaction with Dntf-2
- Switch II
- Interaction with RanGap
- Interaction with exportins
- Interaction with Importin β
- C terminal end interactions

Arrows show positively selected sites inferred using PAML

Tracy et al. In preparation
### Ran-like interactions

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ran H. sapiens</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>Interaction with RCC1</td>
</tr>
<tr>
<td><em>Ran C. familiaris</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>KT Switch I</td>
</tr>
<tr>
<td><em>Ran D. sim</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>KT Switch II</td>
</tr>
<tr>
<td><em>Ran D. sech</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>Interaction with RanGap</td>
</tr>
<tr>
<td><em>Ran D. mel</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>Interaction with exportins</td>
</tr>
<tr>
<td><em>Ran D. erecta</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>Interaction with Importin β</td>
</tr>
<tr>
<td><em>Ran-like D. sim</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>C terminal end interactions</td>
</tr>
<tr>
<td><em>Ran-like D. sech</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>Arrows show positively</td>
</tr>
<tr>
<td><em>Ran-like D. mel</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>selected sites inferred</td>
</tr>
<tr>
<td><em>Ran-like D. erecta</em></td>
<td>MAAGQRPGQVHVLVLCGQGKTTFVHKRITGFEFPTFTVPVVTPFHNPRSGPK 60</td>
<td>using PAML</td>
</tr>
</tbody>
</table>

*Tracy et al.* In preparation
What is meiotic drive?

Meiotic drive (also called segregation distortion) is any process which causes one gametic type to be over-represented in the gametes formed during meiosis, and hence in the next generation.
What is meiotic drive?

Meiotic drive (also called segregation distortion) is any process which causes one gametic type to be over-represented in the gametes formed during meiosis, and hence in the next generation.
SD system in *D. melanogaster*

- Only heterozygous males show distortion
- SD chromosome present in 99% of progeny
SD system in *D. melanogaster*

- Only heterozygous males show distortion
- SD chromosome present in 99% of progeny
SD system in *D. melanogaster*

Figure from Ferree and Barbash, 2007
Segregation distortion in *Dntf-2r* knock down

**Table 1:**

<table>
<thead>
<tr>
<th>Genotype of parent</th>
<th>Type of cross</th>
<th>Phenotype of progeny*</th>
<th>k†</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD-5/cn bw</td>
<td>A</td>
<td>2181 8</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1465 1358</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:**

<table>
<thead>
<tr>
<th>Genotype of parent</th>
<th>Type of cross</th>
<th>Phenotype of progeny*</th>
<th>k†</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD-5/16658</td>
<td>a</td>
<td>1434 419</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>82 68</td>
<td></td>
</tr>
</tbody>
</table>

- **k** values are the proportion of SD-bearing progeny among the total offspring.
- 0.99 is significantly different from 0.77 (Fisher’s Exact Test P<10⁻⁷)
- 82/68 is not significantly different from 75/75 ($X^2=1.3067$; d.f.=1; P=0.2530)
- No obvious fertility effects of *Dntf-2r* knockout
Why recurrent recruitment of Ntf-2 and Ran?

Hypothesis III: Meiotic drive role
– Selfish meiotic drive systems evolve all the time and these genes might have a role as drivers or suppressors
– I also like to speculate that they might also have an interplay with sexual antagonism

Supported by loss of new retrogenes, loss of functions of the new retrogenes, and lack of infertility effects of null alleles of *Dntf-2r* (Tracy *et al*. In preparation) and high turnover of species restricted genes with male biased expression (Zhang *et al*. 2007)
Y chromosome and gene duplication

At the end the non-recombining region of the Y chromosome is left with very few genes that are male-specific.
The rise and fall of the human Y chromosome
Marshall-Graves 2002
The sequencing of the Y chromosome: rethinking the rotting Y chromosome
David Page

Associate Director of Science, Whitehead Institute
Professor of Biology, MIT
Investigator of the Howard Hughes Medical Institute
Determining the sequence of the human Y chromosome presented a daunting challenge. But the task is now done and the secrets revealed justify the effort.

Skaletsky et al. 2003; Willard 2003
Eight palindromes comprise one-quarter of the euchromatic region of the Y chromosome Male Specific Region (MSY; previously named NRY or non-recombining region)

Figure 3  The Y chromosome is highly repetitive. A section of the Y chromosome that David Page studies, called AZFc (for azospermia factor c), consists of DNA sequences that read the same in either direction, an organization that can lead to instability as well as provide a mechanism to evolve new alleles. Other parts of the chromosome house similar repeats. Matching colors in this depiction represent identical sequences. Same-color arrows that point in opposite directions indicate inverted repeats, similar to palindromes in the English language.

Skaletsky et al. 2003; Rozen et al. 2003
This came as an incredible surprise. There are no regions in the genome organized this way.

Skaletsky et al. 2003; Rozen et al. 2003
Abundant gene conversion between the arms of palindromes in human and ape Y chromosomes

Rozen et al. 2003
Abundant gene conversion between the arms of palindromes in human and ape Y chromosomes

Rozen et al. 2003
Concerted evolution between the arms of palindromes in human and ape Y chromosomes

Rozen et al. 2003
Sex chromosomes evolution: a genomic make over
A lot of aspects of molecular evolution can be exemplified with sex chromosomes evolution:

1. Whole chromosome gene expression changes/ Dosage compensation.
2. Chromosome degeneration and specialization
3. Effects of the lack of recombination
4. Chromosome reorganizations: inversions and translocations
5. Duplication into the Y, in the Y, out of the X and into the X
6. Accumulation of TE elements
7. Positive selection
Current and future directions

• How do retrogenes recruit new promoters?
• What are the functions of the young retrogenes in *Drosophila*? Are some of them involved in meiotic drive? What about in other organisms?
• How different are the mitochondria and the proteasome during spermatogenesis?
• When does the export begin?
• Does it hold for other types of duplicates?
• Is there a pattern of duplication in other organisms?
• Is the X inactivated in *Drosophila*? Cytological approach
• Transposable element protein domestication
• Duplicated genes at the breakpoints of inversions
Lab members

Research Assistant:
- Javier Rio

Postdocs:
- Claudio Casola
- Miguel Gallach
- Jorge E. Quezada-Diaz

Graduate students:
- Yongsheng Bai
- Susana Domingues
- Aditi Bhardwaj
- Mansi Motiwale
- Taniya Muliyil
- Mehran Sorourian
- Daniel Szalay
- Charles Tracy

Undergraduate students
- Sarah Cole
- Jamie Dunlap
- Anton Mare
- Peter Pelache

Support: R01 from NIH and UTA (REP grant and startup funds)
Collaborators

- Chitra Chandrasekaran
- Ying Chen
- Shawn Christensen
- Isabelle Dupanloup
- J.J. Emerson
- Cédric Feschotte
- Henrik Kaessmann
- Manyuan Long
- Alexandre Reymond
- Alfredo Ruiz
- Kevin Thornton

University of Chicago, University of Lausanne, Wesleyan University and UTA