

# Modeling and Displaying\* Synteny

A cautionary tale

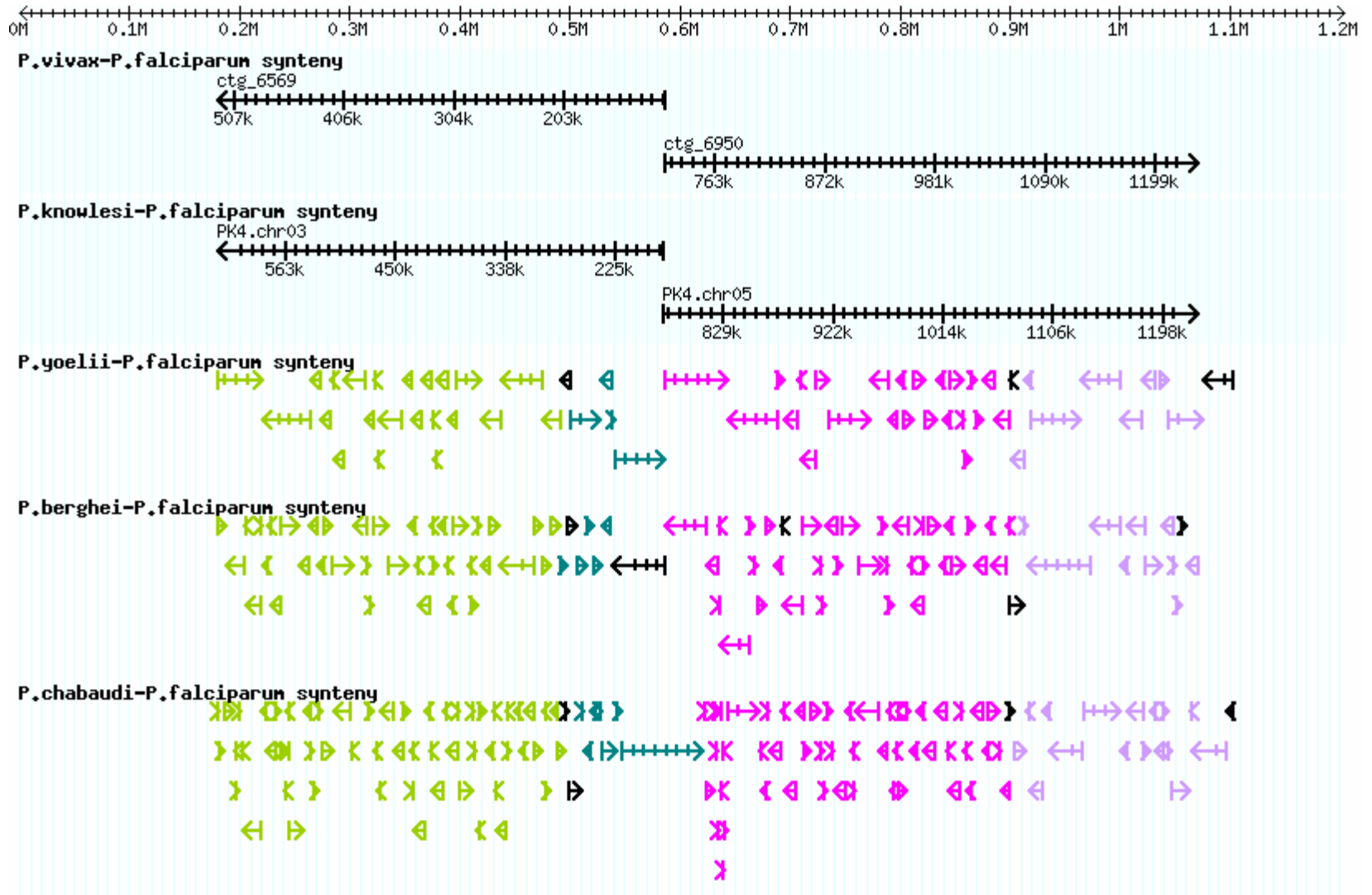
\* in gbrowse

# Overview

- Getting familiar w/ SynView
- Modeling the data
- Serving the data to gbrowse/SynView
- The ***PERILS*** of artifacts

# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

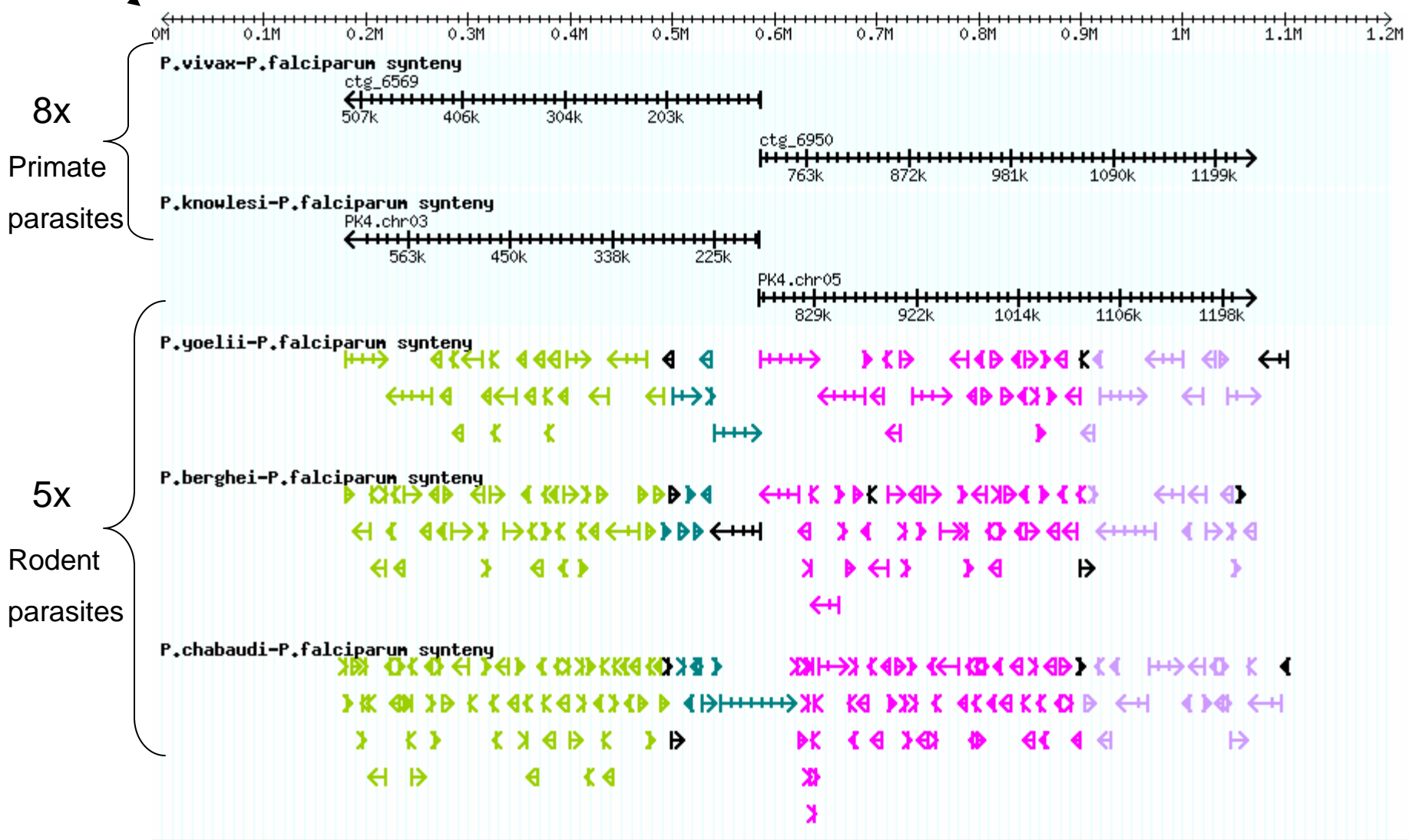
-- chromosome view --



# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

-- chromosome view --

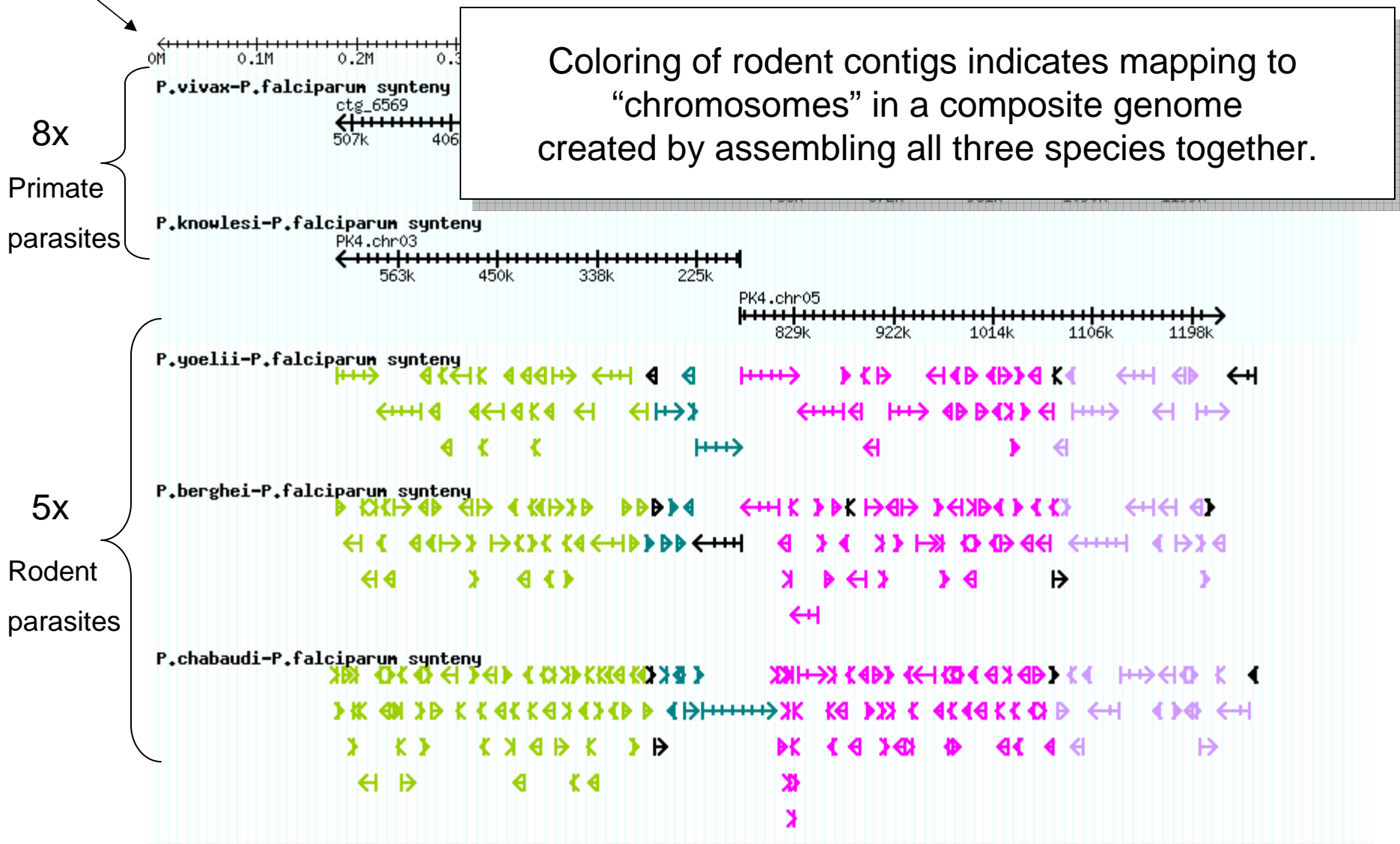
P.f. (human parasite) is finished



# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

-- chromosome view --

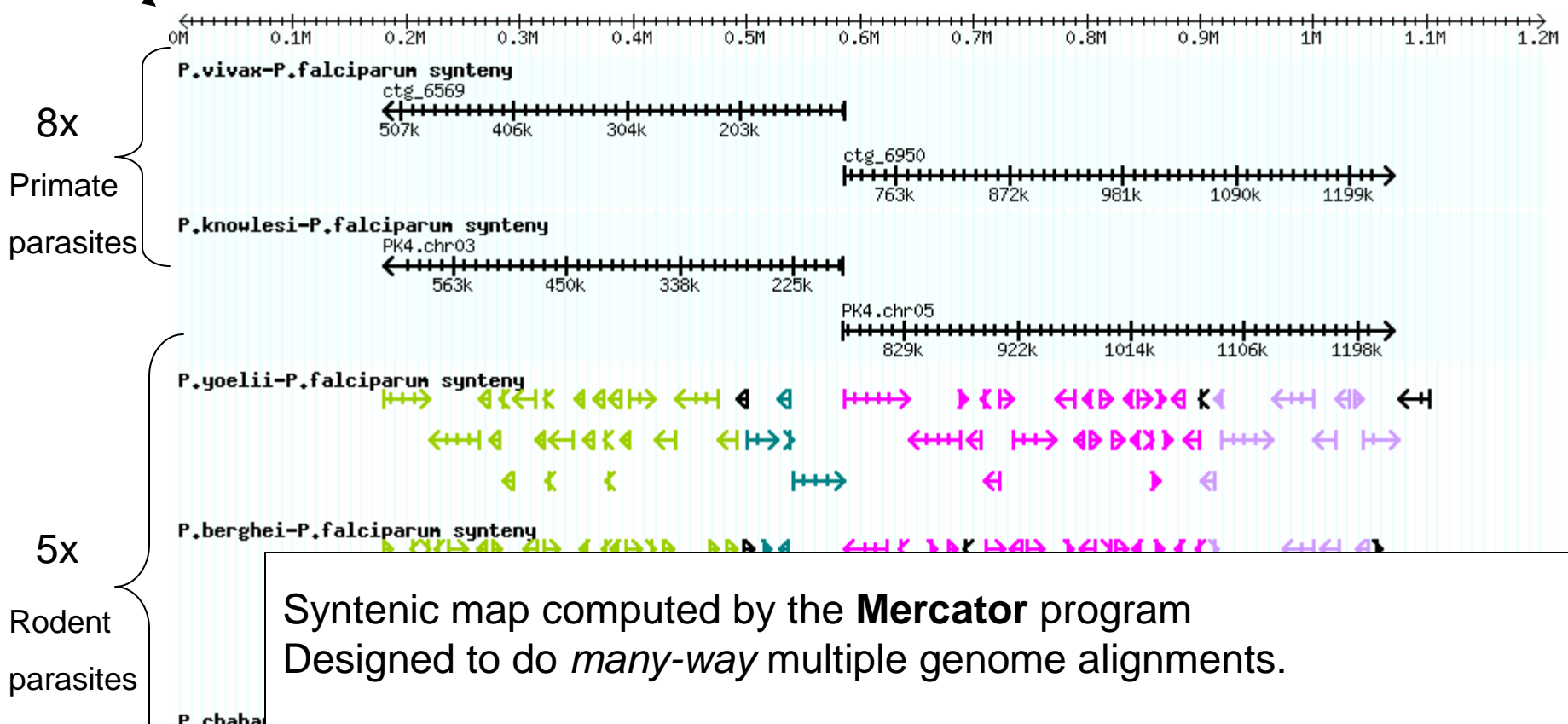
P.f. (human parasite) is finished



# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

-- chromosome view --

P.f. (human parasite) is finished

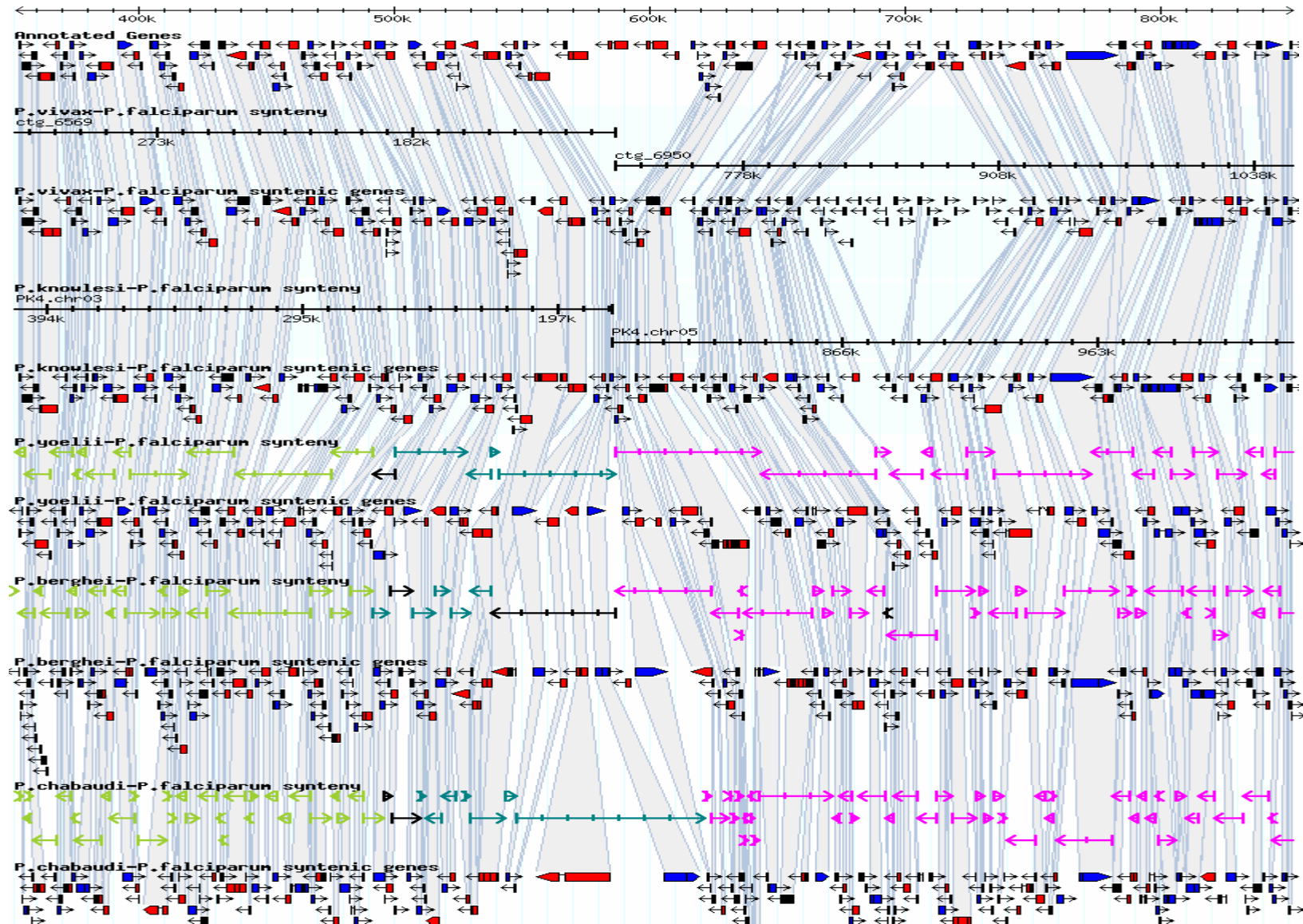


Syntenic map computed by the **Mercator** program  
 Designed to do *many-way* multiple genome alignments.

[C. Dewey](#) (2007) Aligning multiple whole genomes with Mercator and MAVID.  
 In N. Bergman, editor, [Comparative Genomics](#), volume 395 of  
[Methods in Molecular Biology](#). [Humana Press](#).

# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

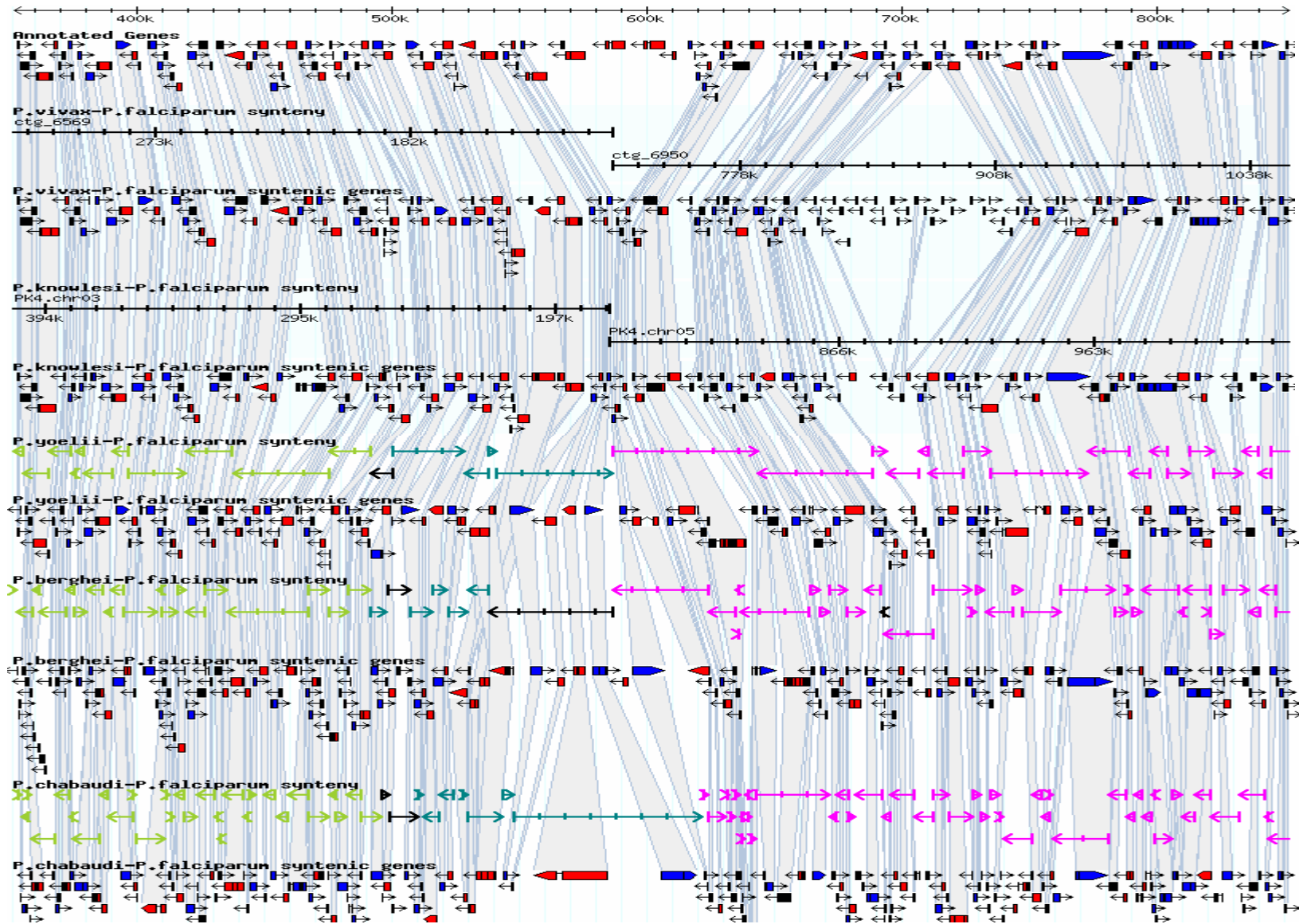
-- gene view --



# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

-- gene view --

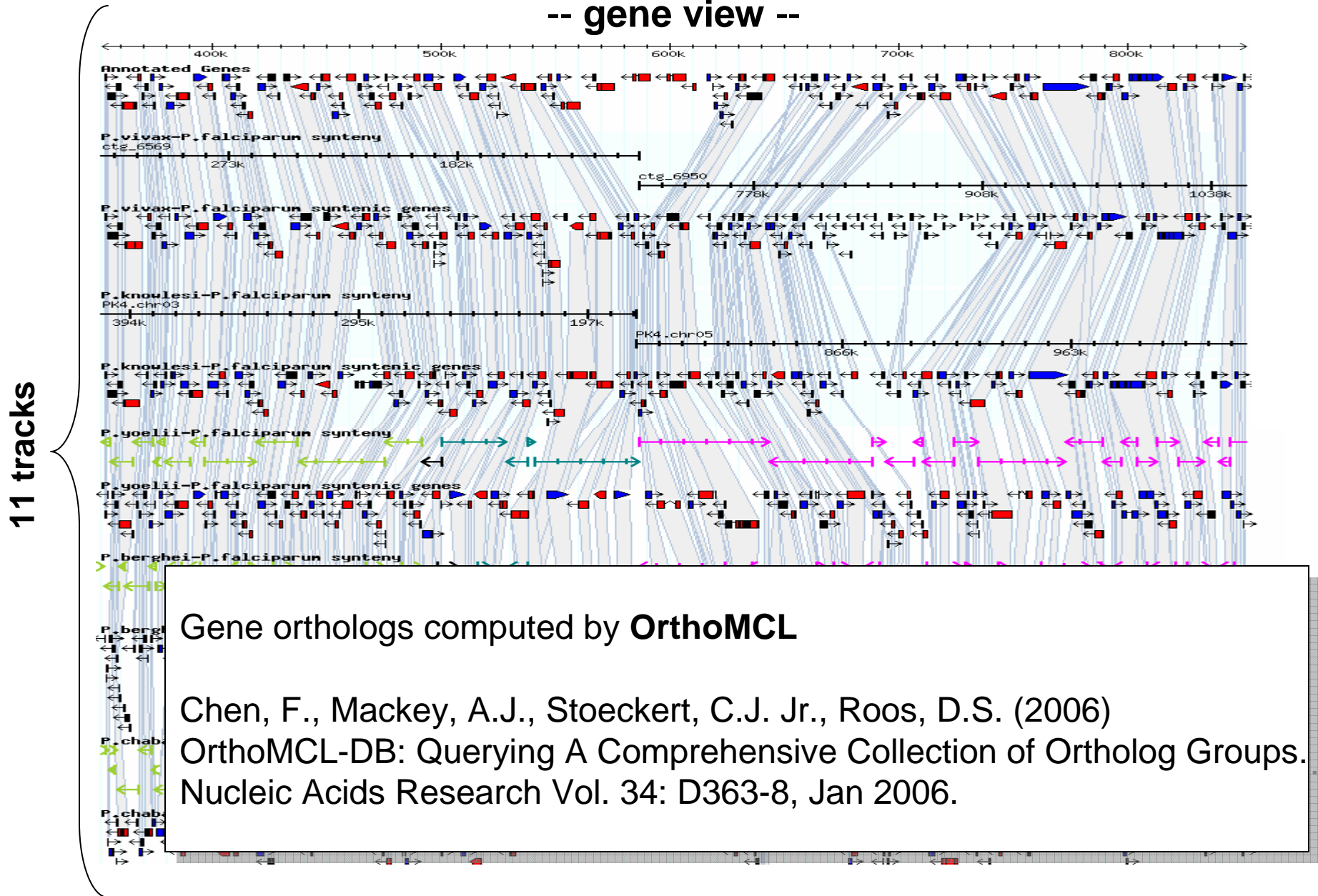
11 tracks





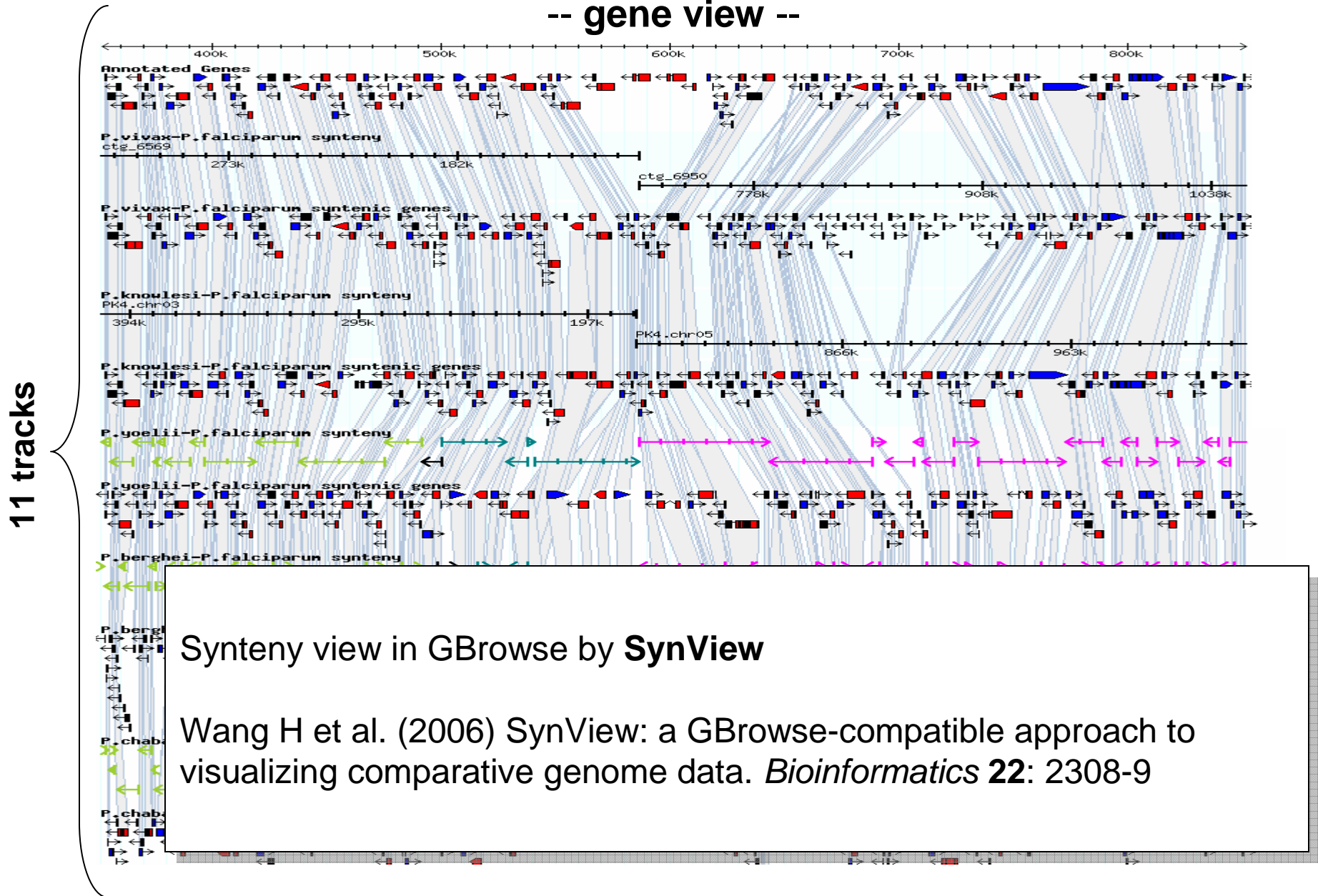
# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

-- gene view --



# Synteny of 5 *Plasmodium* species against *P. falciparum* chr 4 (reference)

-- gene view --



Synteny view in GBrowse by **SynView**

Wang H et al. (2006) SynView: a GBrowse-compatible approach to visualizing comparative genome data. *Bioinformatics* **22**: 2308-9

# Computing the display

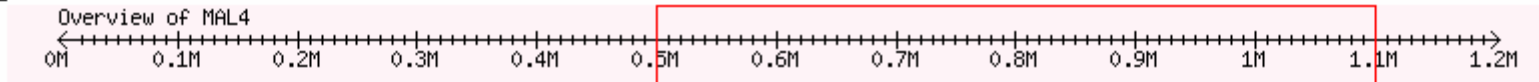
**Landmark or Region:** mal4:500000-1100000  Search

**Reports & Analysis:** Annotate Restriction Sites  Configure...

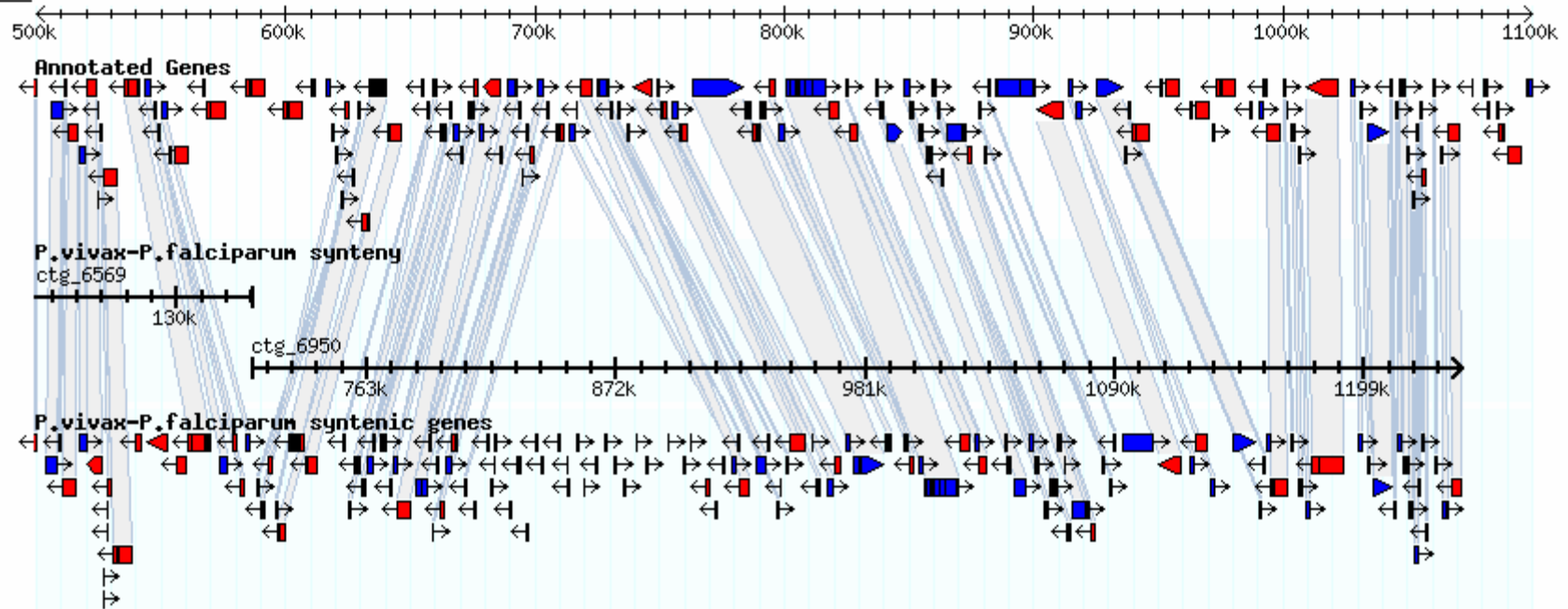
**Data Source:** Genome Sequence

**Scroll/Zoom:**     Show 600 kbp     Flip

## Overview



## Details



# Computing the display

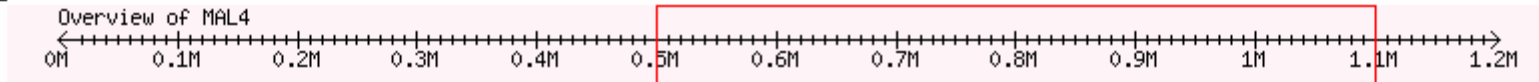
**Landmark or Region:** mal4:500000-1100000

**Reports & Analysis:** Annotate Restriction Sites

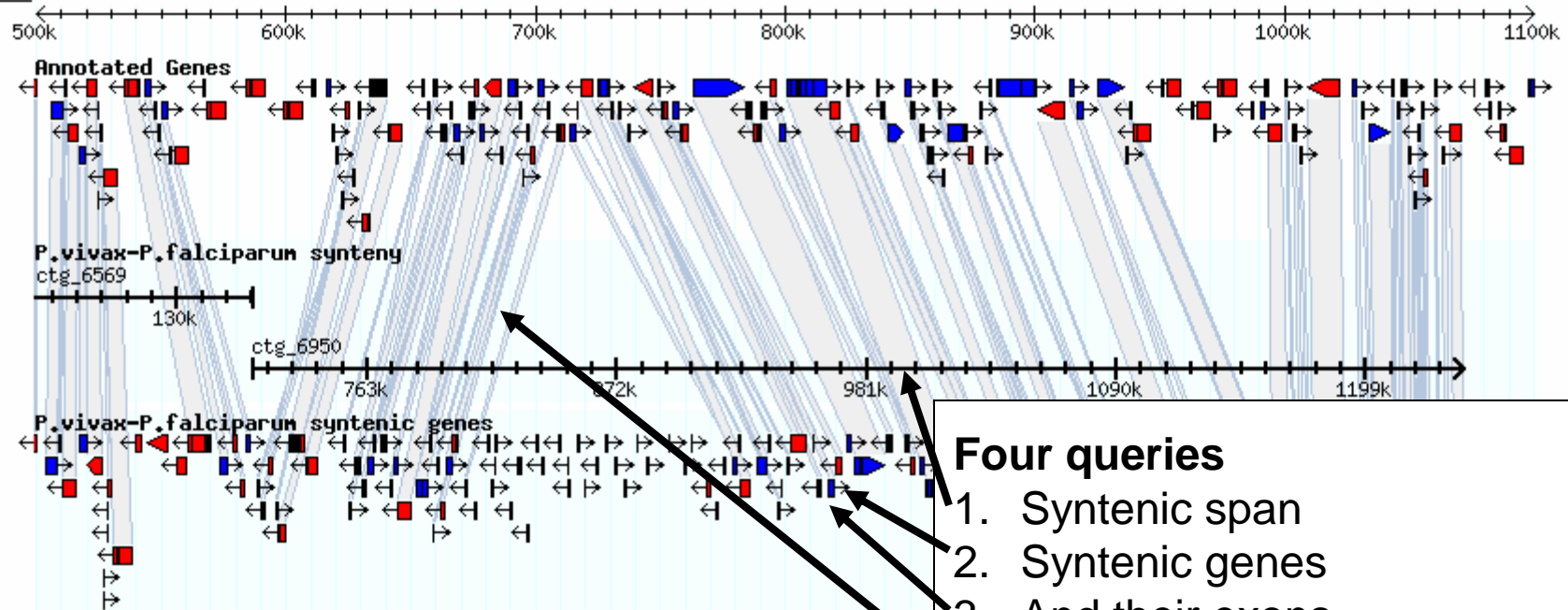
**Data Source:** Genome Sequence

**Scroll/Zoom:**   Show 600 kbp    Flip

## Overview



## Details



### Four queries

1. Syntenic span
2. Syntenic genes
3. And their exons
4. Gene orthology

## *P. falciparum* Chr 4 (reference)

Landmark or Region:

mal4:500000-1100000

Search

Reports & Analysis:

Annotate Restriction Sites

Configure...

Go

Data Source

Genome Sequence

Scroll/Zoom: <<< - + >>>

Show 600 kbp

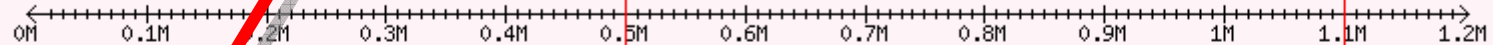
+

>>>

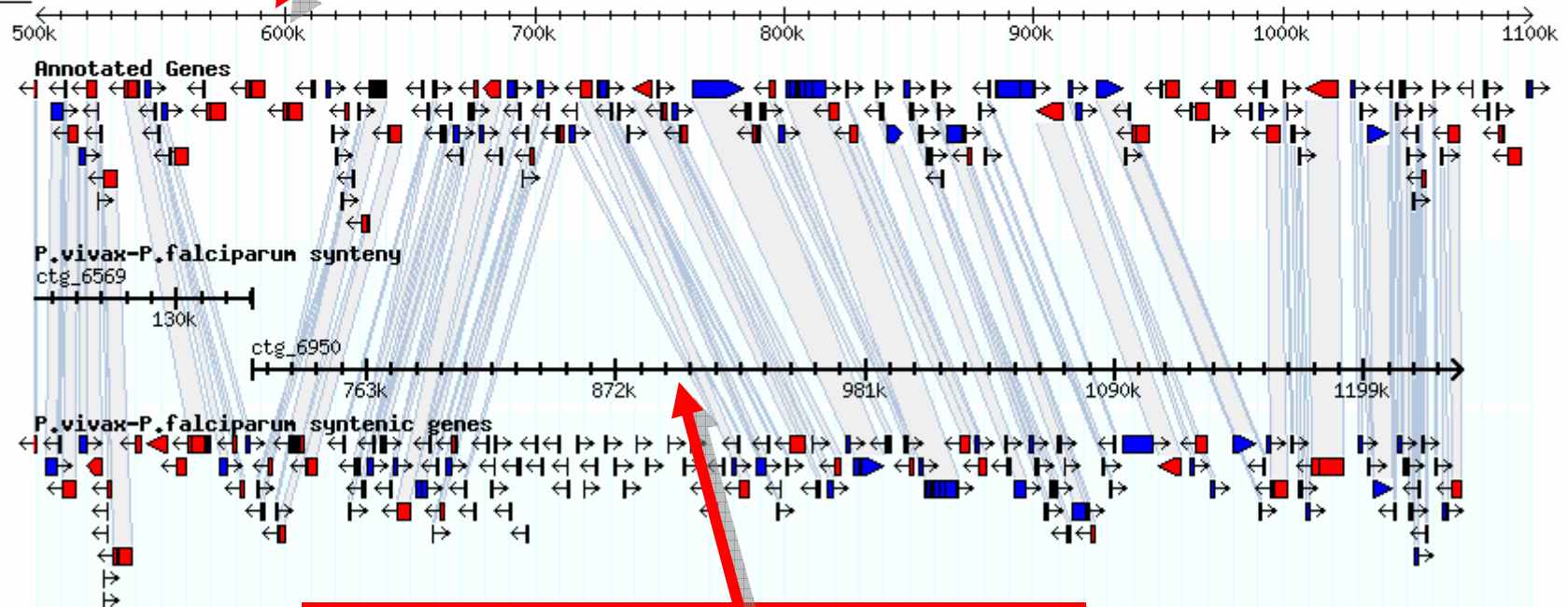
Flip

Overview

Overview of MAL4



Details



## *P. vivax* contig (syntenic)

## *P. falciparum* Chr 4 (reference)

Landmark or Region:

mal4:500000-1100000

Search

Reports & Analysis:

Annotate Restriction Sites

Configure...

Go

Data Source

Genome Sequence

Scroll/Zoom:

<<<

—

Show 600 kbp

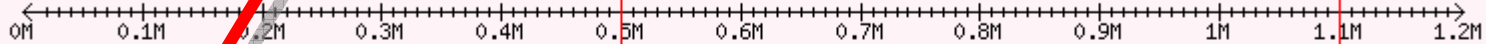
+

>>>

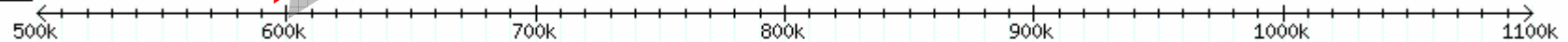
Flip

Overview

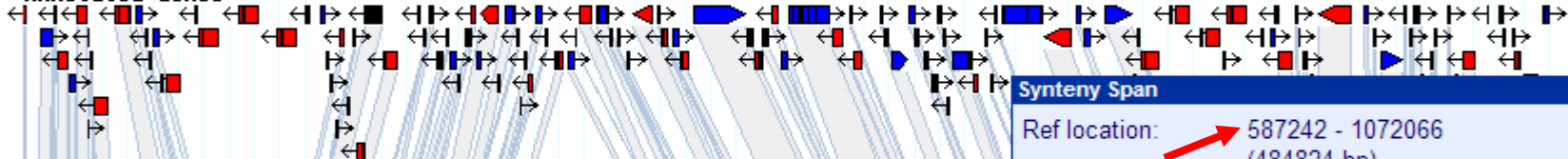
Overview of MAL4



Details



Annotated Genes



*P. vivax*-*P. falciparum* synteny

ctg\_6569

130k

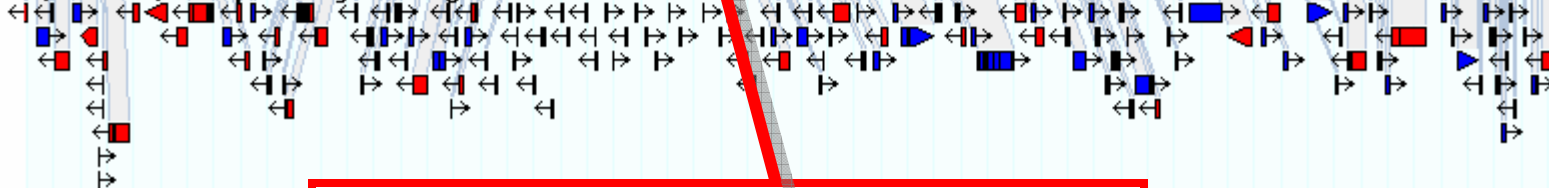
ctg\_6950

763k

872k

981k

*P. vivax*-*P. falciparum* syntenic genes



Synteny Span

Ref location: 587242 - 1072066  
(484824 bp)

Syn location: 713644 - 1242205  
(528561 bp)

Reversed: no

Length of syn  
contig: 1370936

## *P. vivax* contig (syntenic)

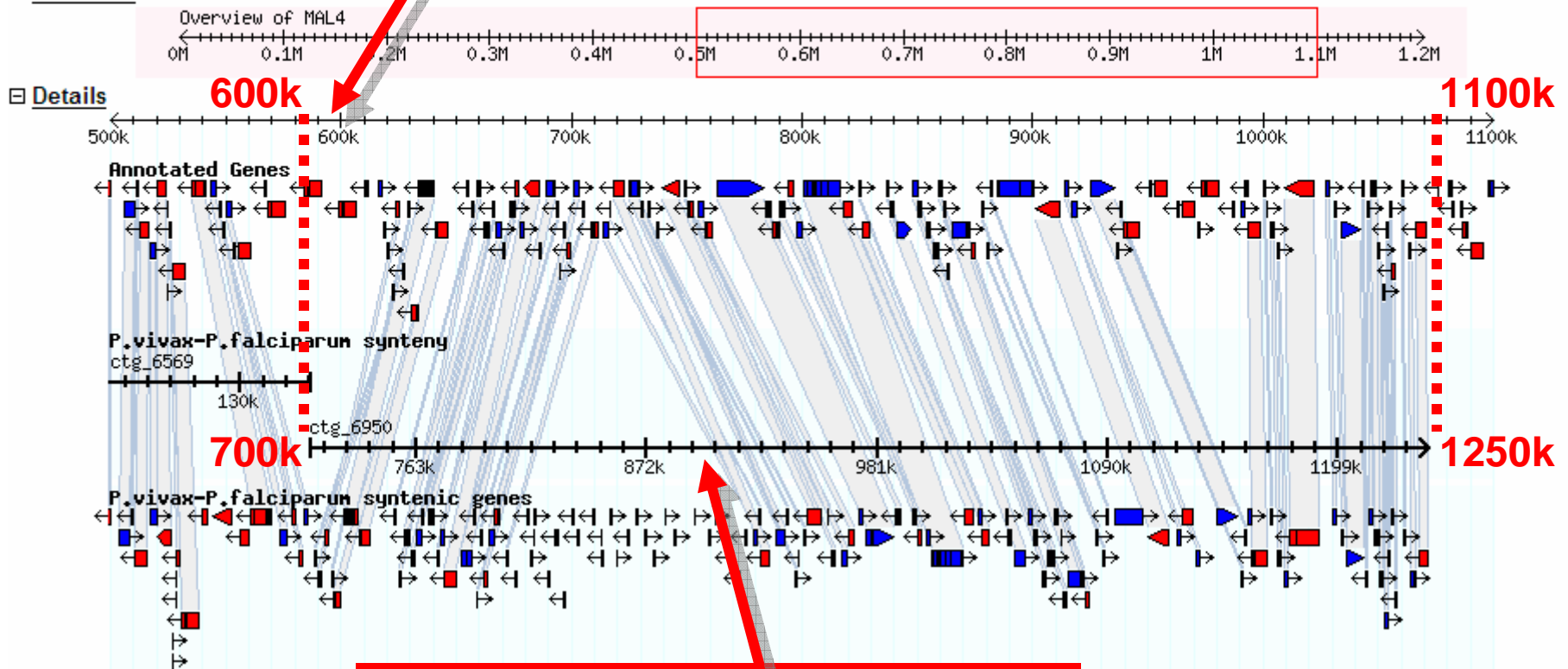
***P. falciparum* Chr 4 (reference)**

Landmark or Region:

Data Source: Genome Sequence

Reports & Analysis:

Scroll/Zoom:         Flip



***P. vivax* contig (syntenic)**

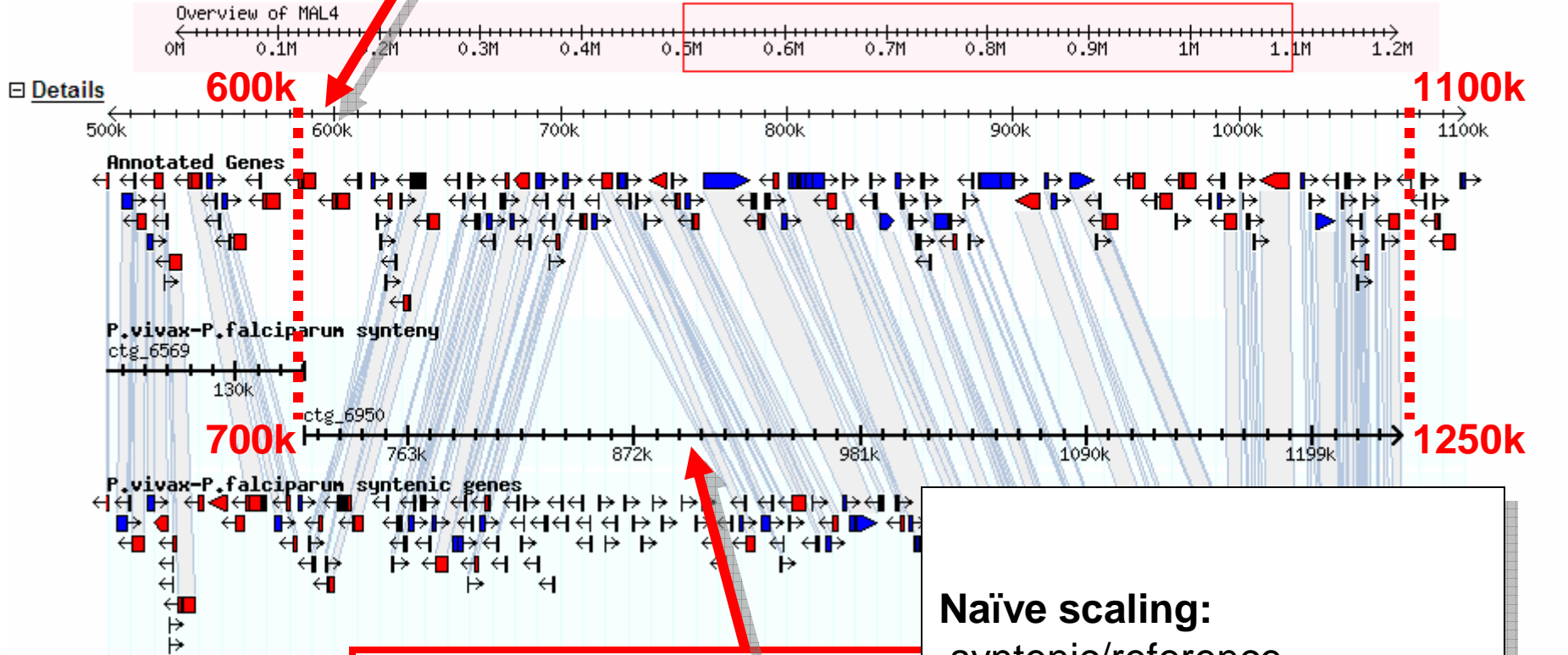
***P. falciparum* Chr 4 (reference)**

Landmark or Region:

Data Source: Genome Sequence

Reports & Analysis:

Scroll/Zoom:         Flip



***P. vivax* contig (syntenic)**



Landmark or Region:

mal4:500000-1100000 Search

Reports & Analysis:

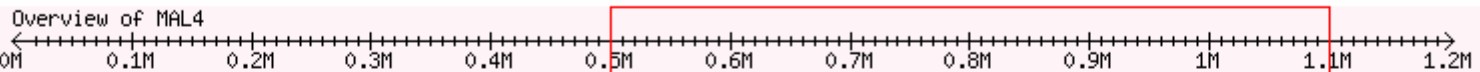
Annotate Restriction Sites Configure... Go

Data Source

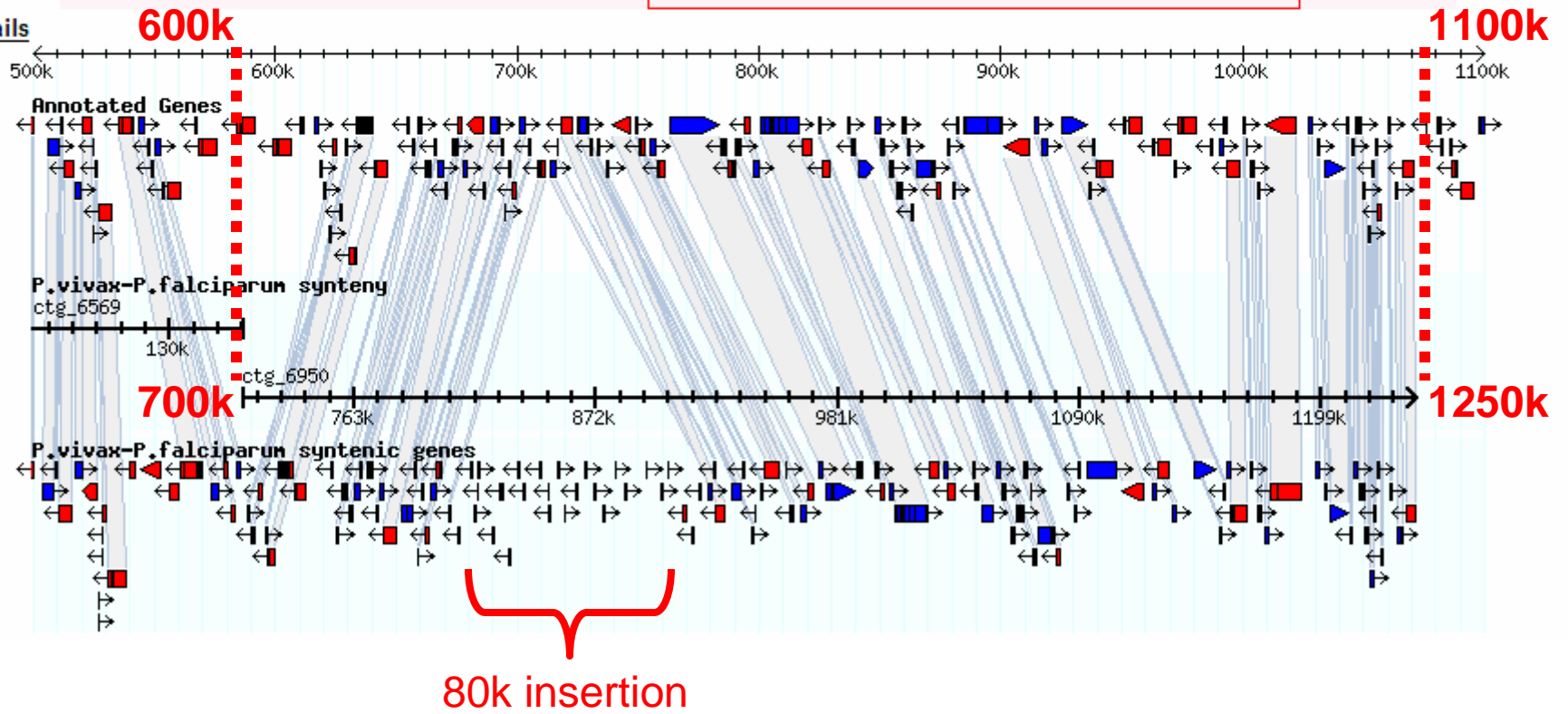
Genome Sequence

Scroll/Zoom: <<< Show 600 kbp + >>> Flip

Overview



Details



Landmark or Region:

mal4:500000-1100000 Search

Reports & Analysis:

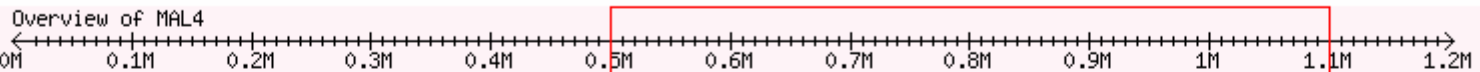
Annotate Restriction Sites Configure... Go

Data Source

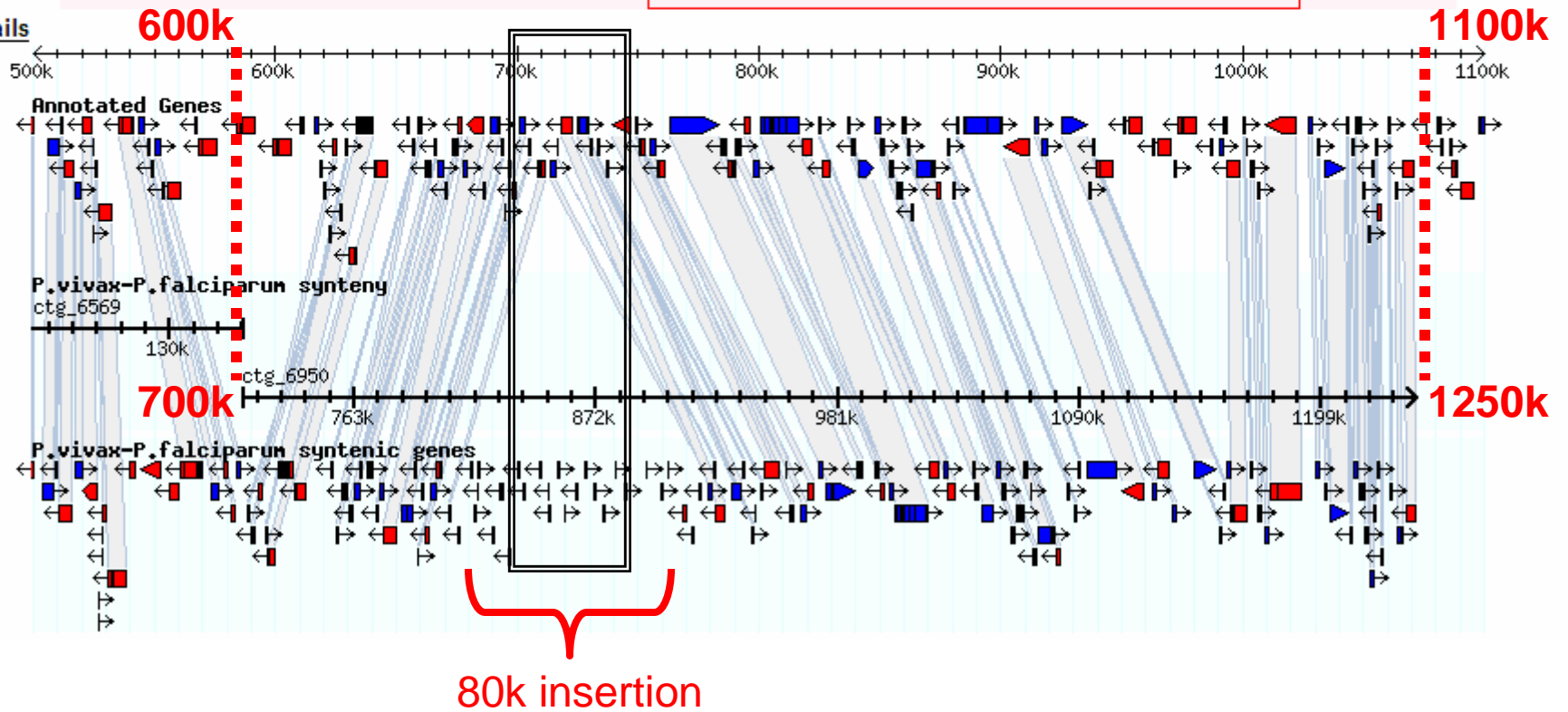
Genome Sequence

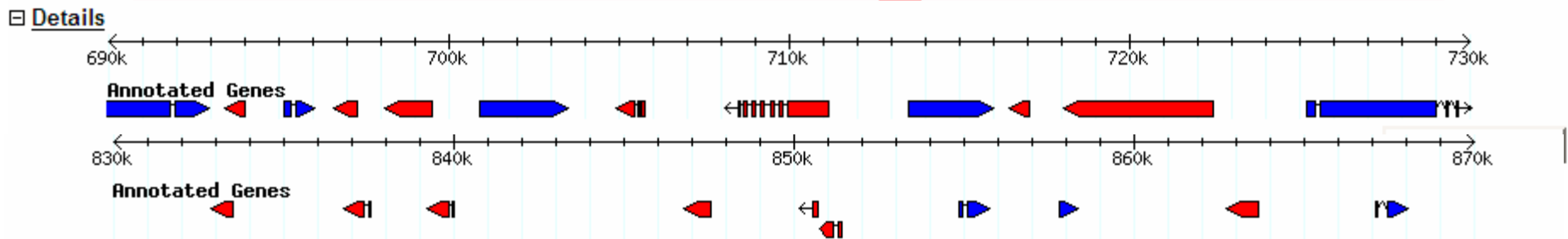
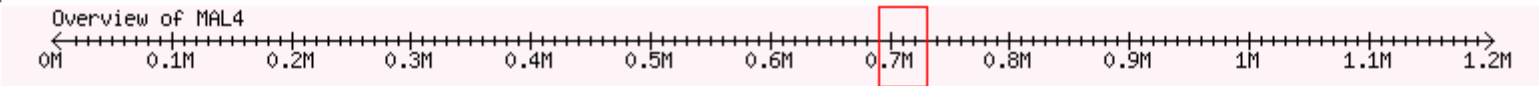
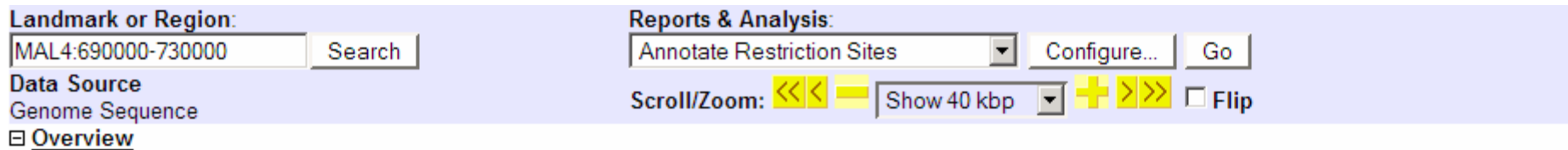
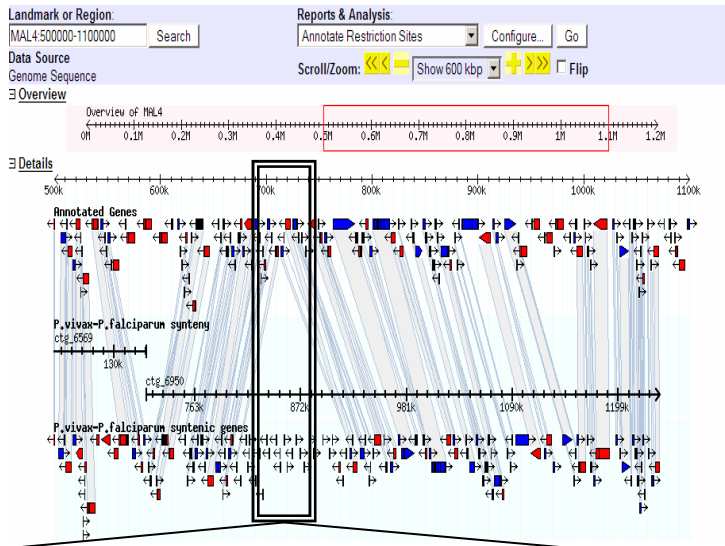
Scroll/Zoom: <<< Show 600 kbp + >>> Flip

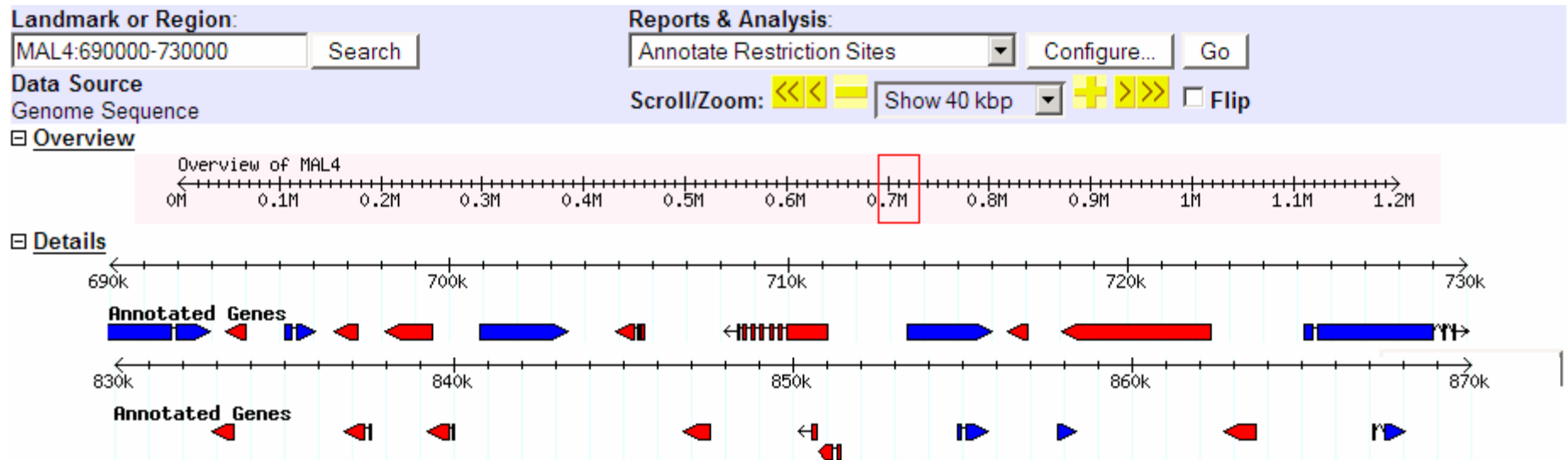
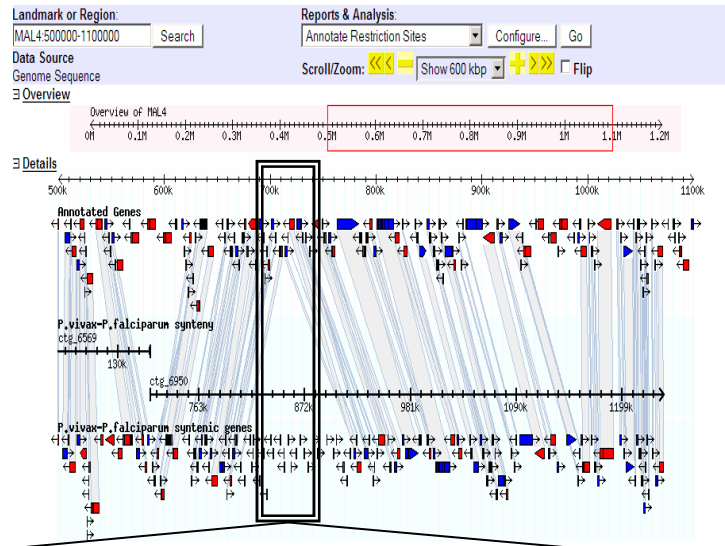
Overview



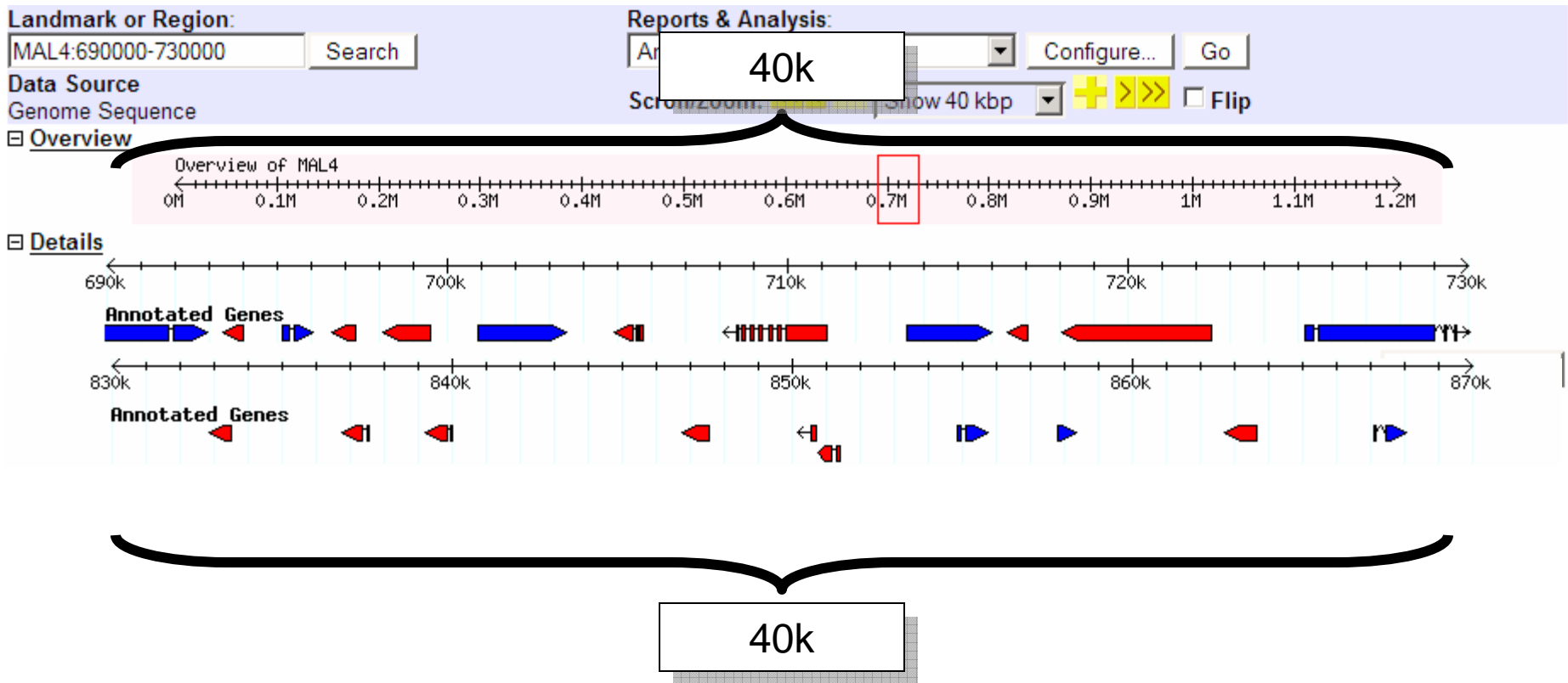
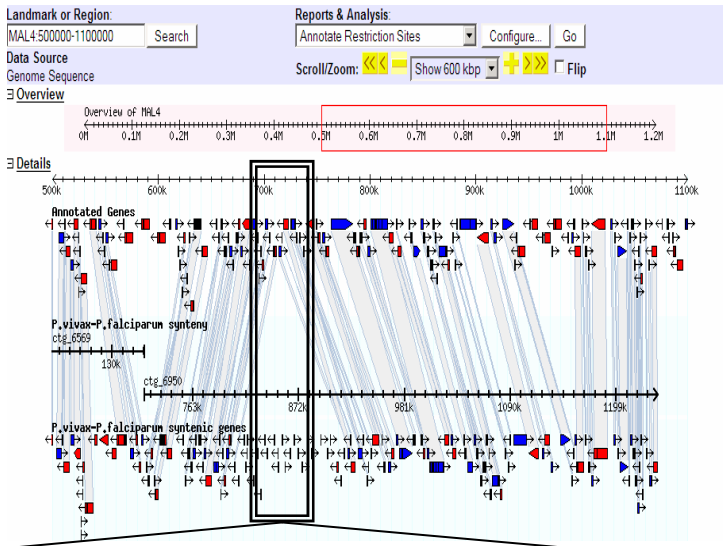
Details

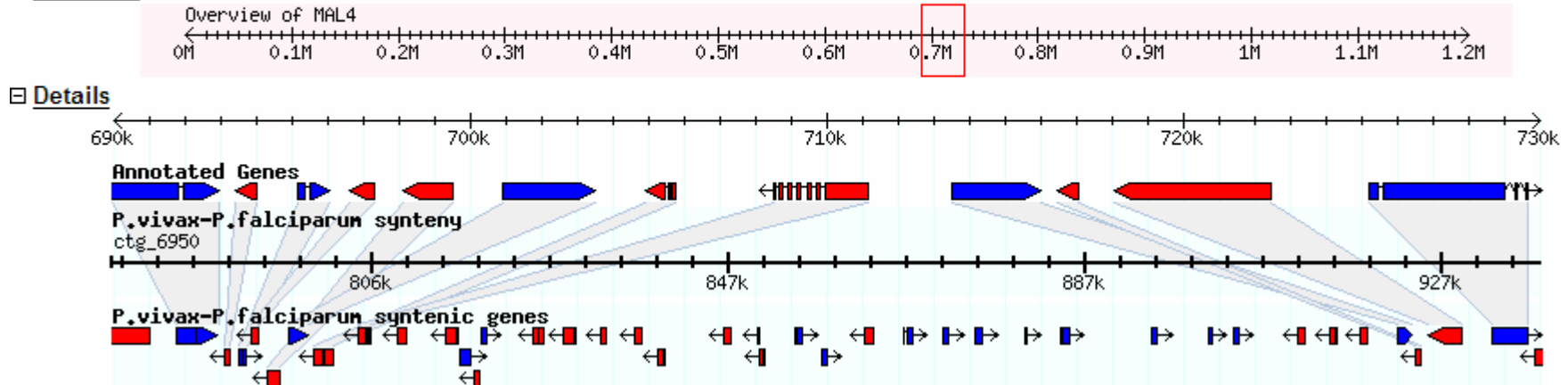
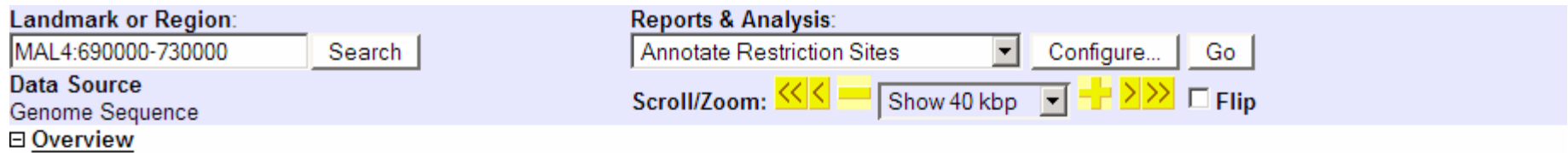
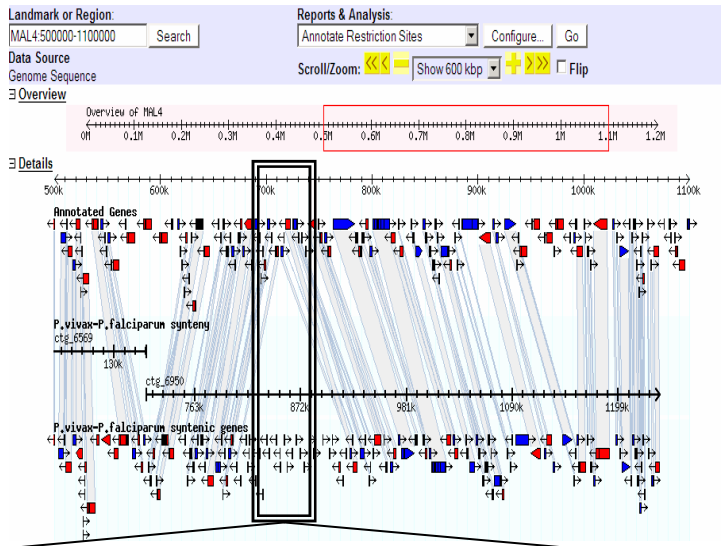


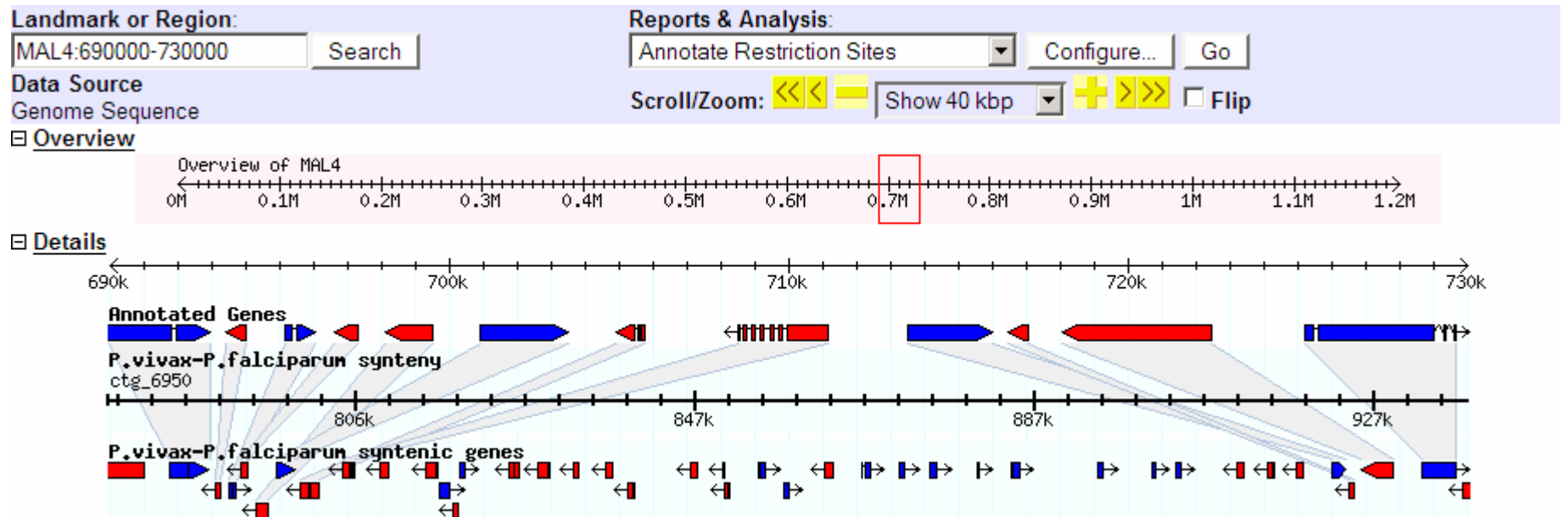
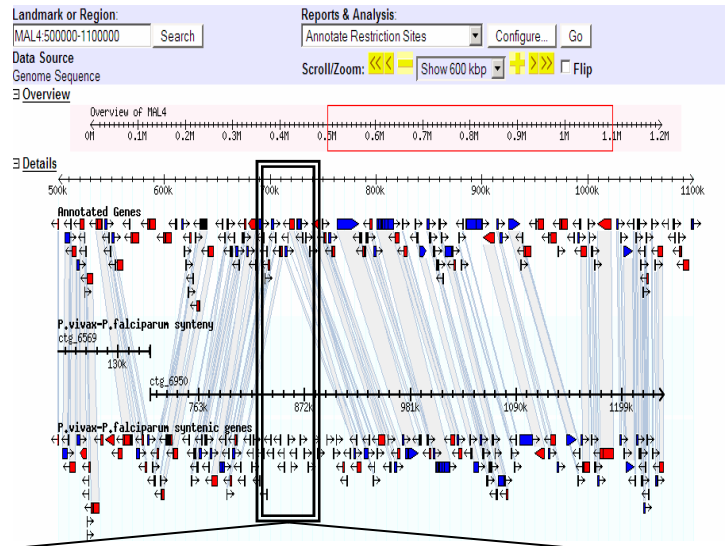




**Artifact!** It appears to the user that the Pf genes have no orthologs!  
 ... they do, but they are **off-screen!**





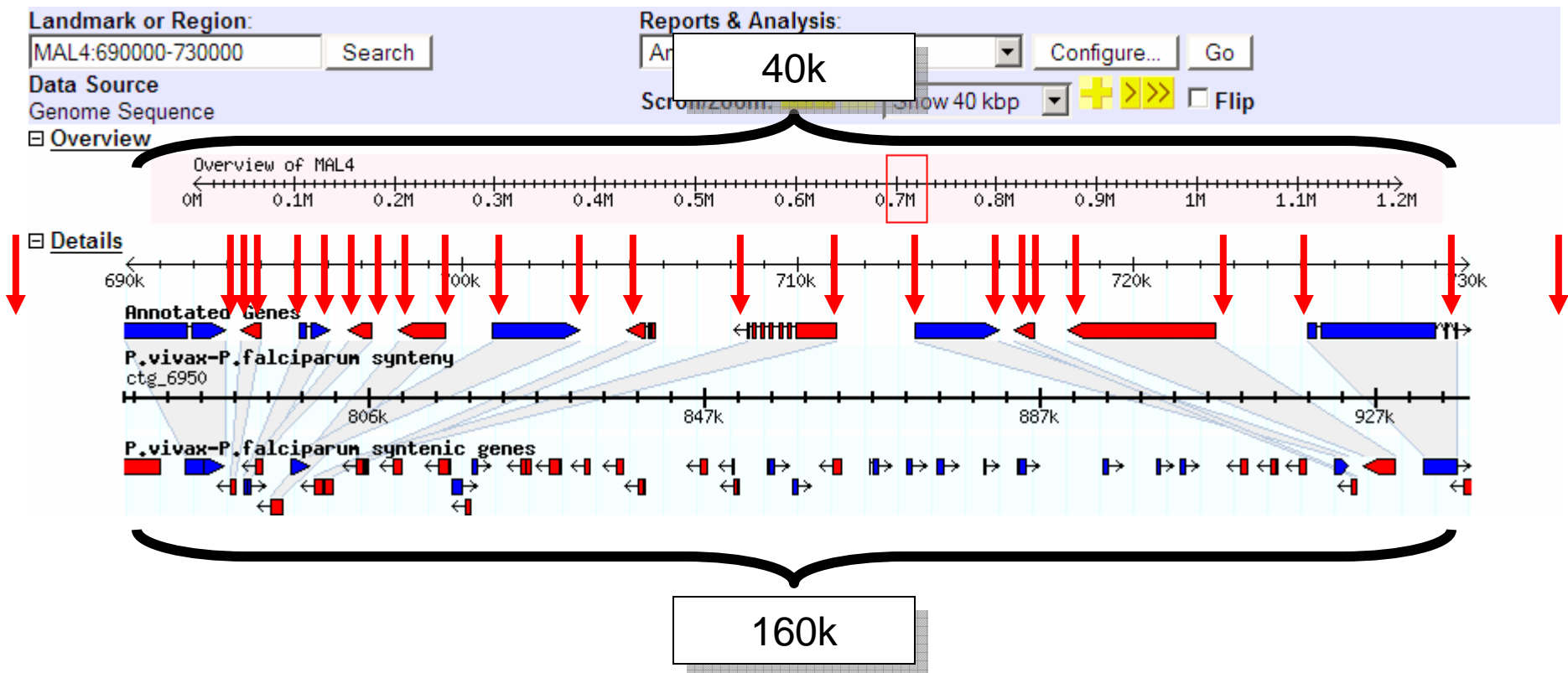


Solution: *scale locally*

Define anchors along reference:

- start of a syntenic segment
- start of genes
- end of genes
- end of a syntenic segment

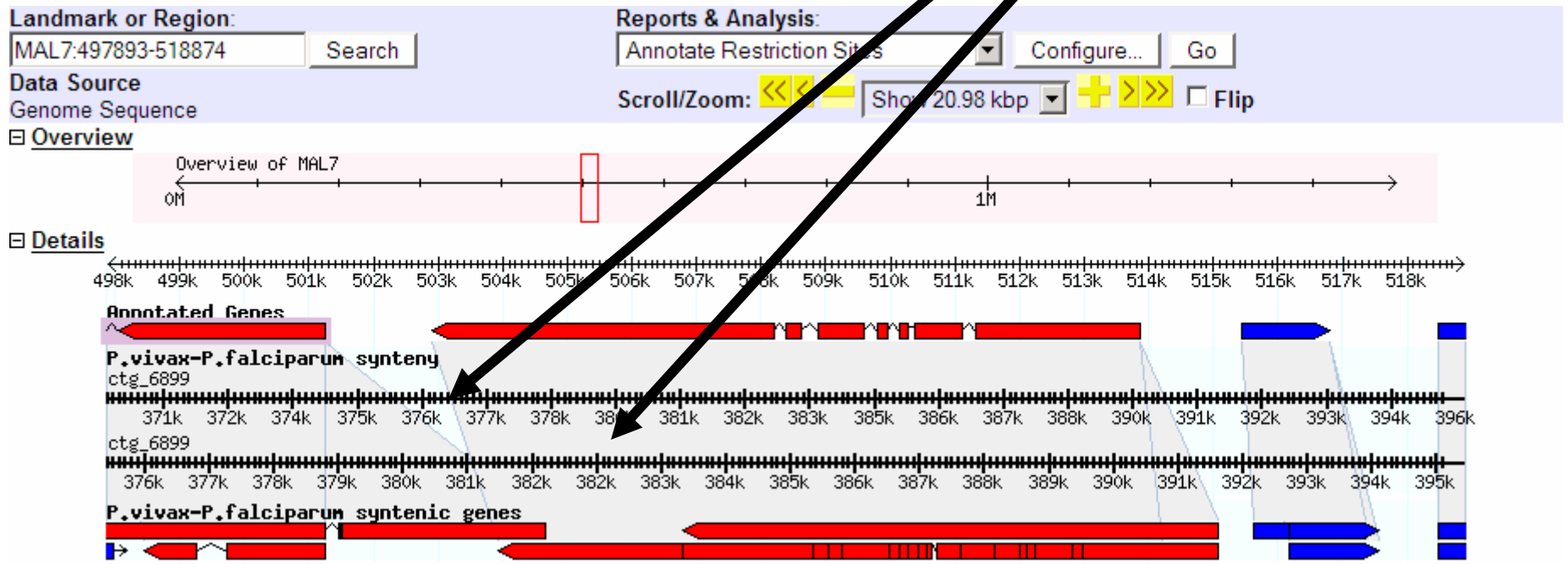
Use left- and right-most visible anchors as scaling bounds





Another artifact causes this **bug**

**Contig ctg\_6899 shown in duplicate**



The source of the problem is a gene model discrepancy, leading to **one** Pf gene with **two** Pv orthologs.

This makes ambiguous anchors.

Landmark or Region:

MAL7:478384..538383 Search

Data Source

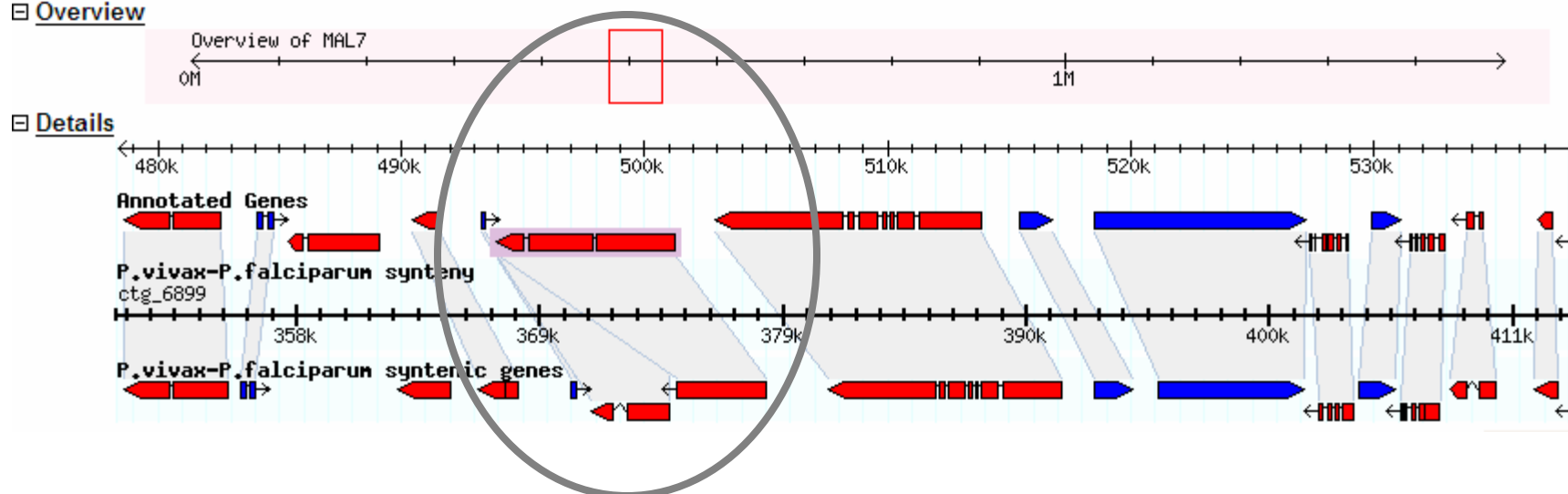
Genome Sequence

Overview

Reports & Analysis:

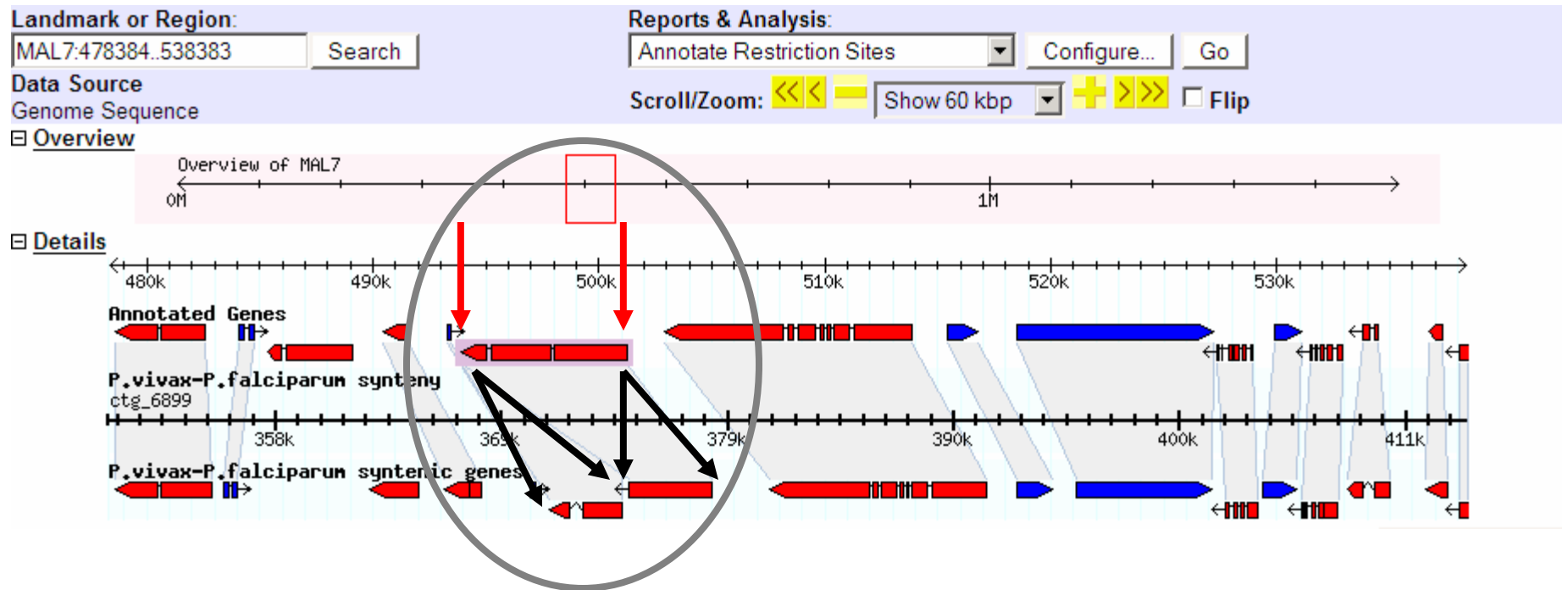
Annotate Restriction Sites Configure... Go

Scroll/Zoom: <<< - Show 60 kbp + >>> Flip



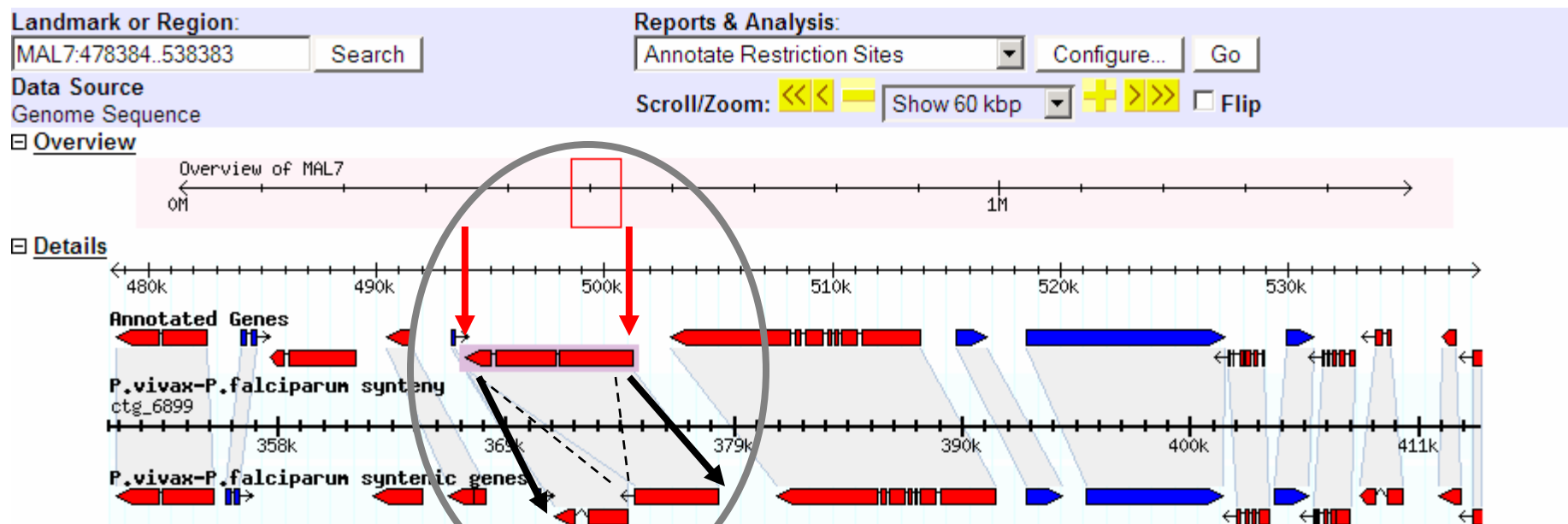
The source of the problem is a gene model discrepancy, leading to **one** Pf gene with **two** Pv orthologs.

This makes ambiguous anchors.



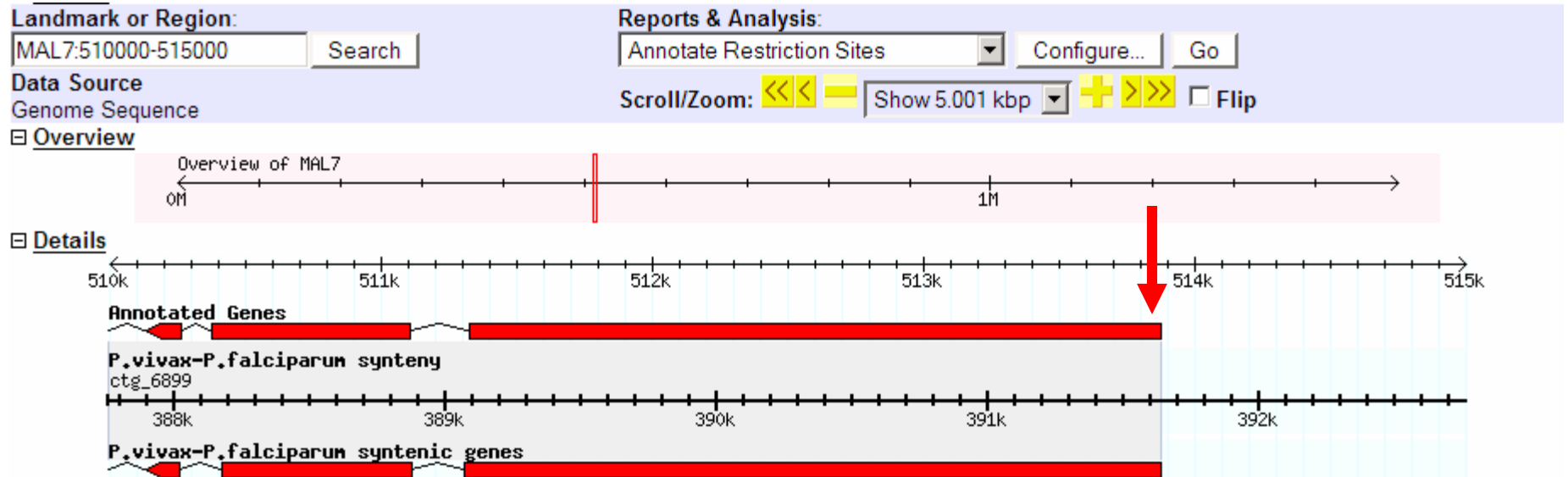
The source of the problem is a gene model discrepancy, leading to **one** Pf gene with **two** Pv orthologs.

This makes ambiguous anchors.



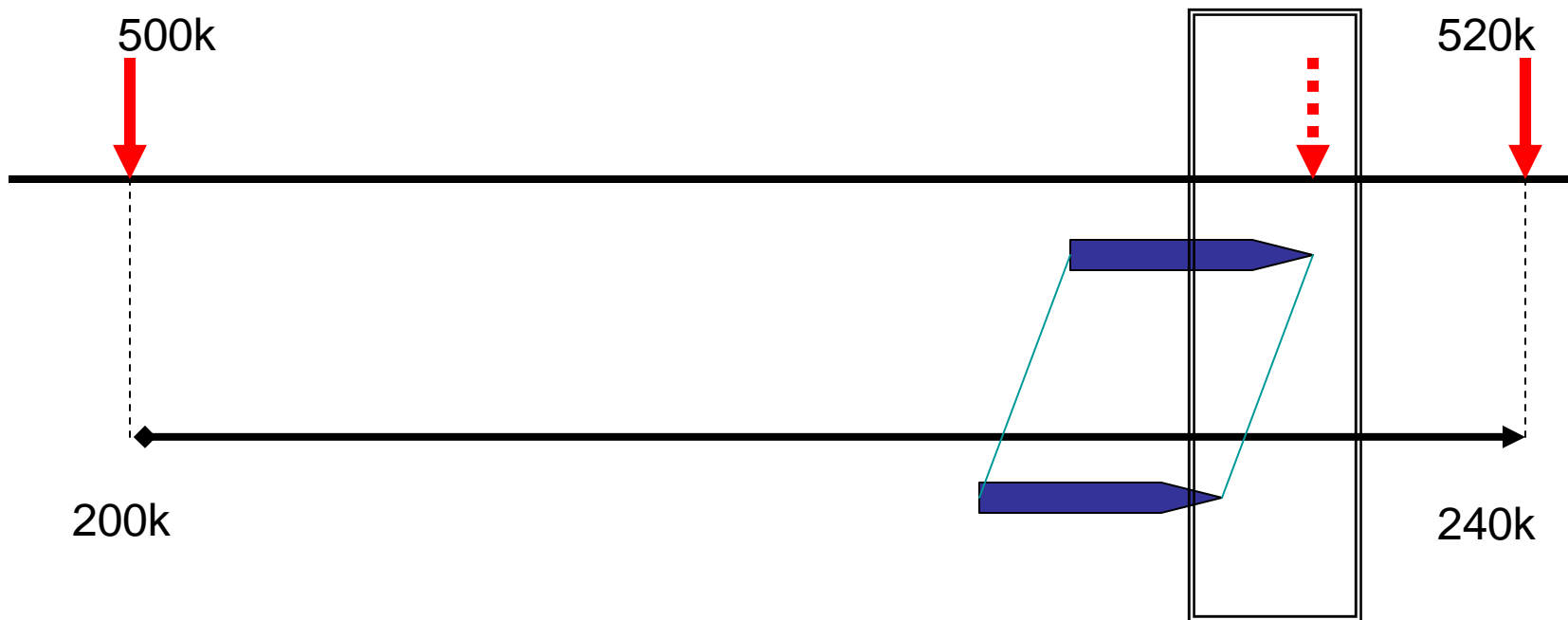
Solution: lose **internal** anchors

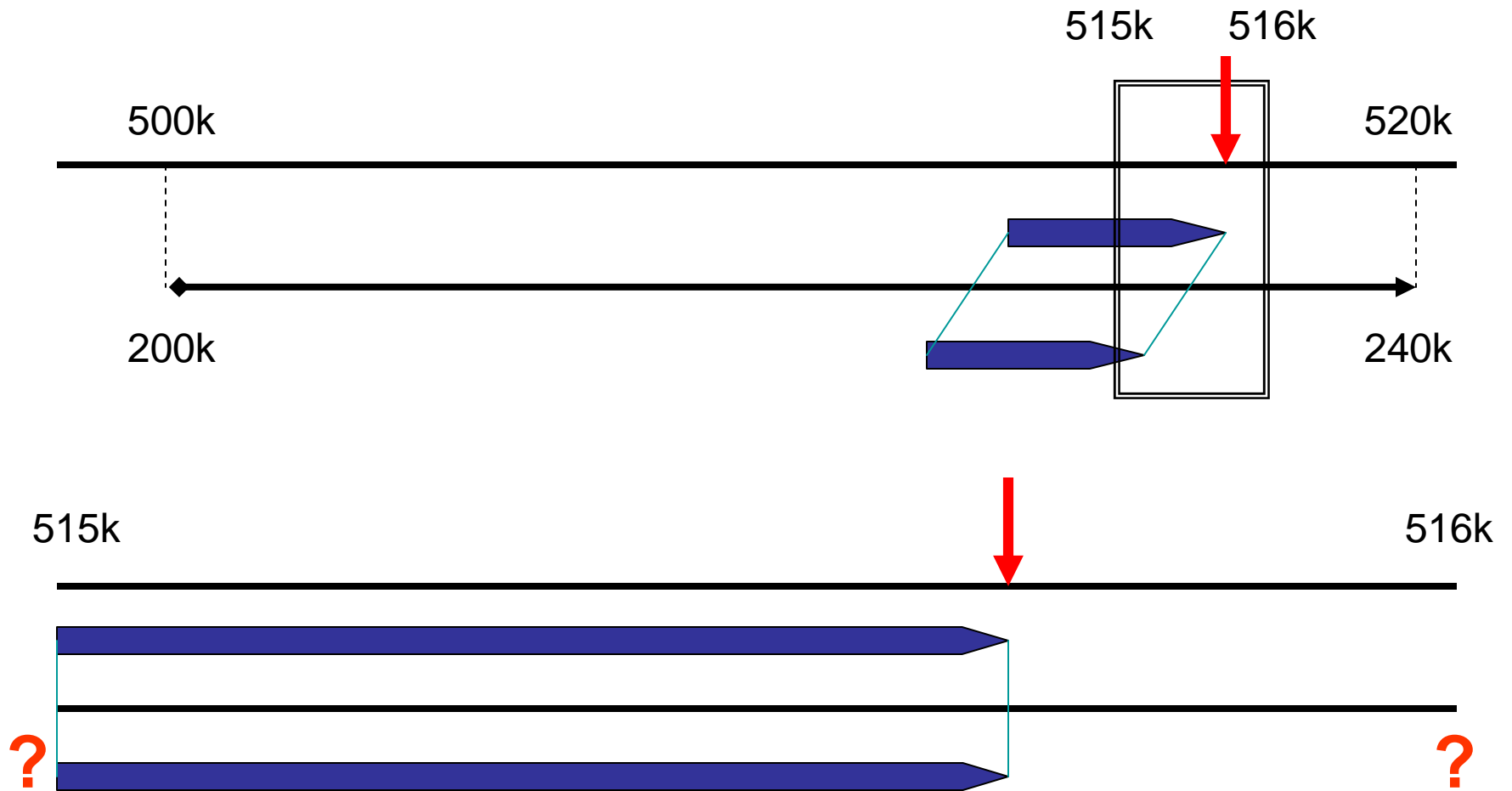
Another problem arises when there is only **one anchor** on screen



Synteny between a  
chromosome and contig

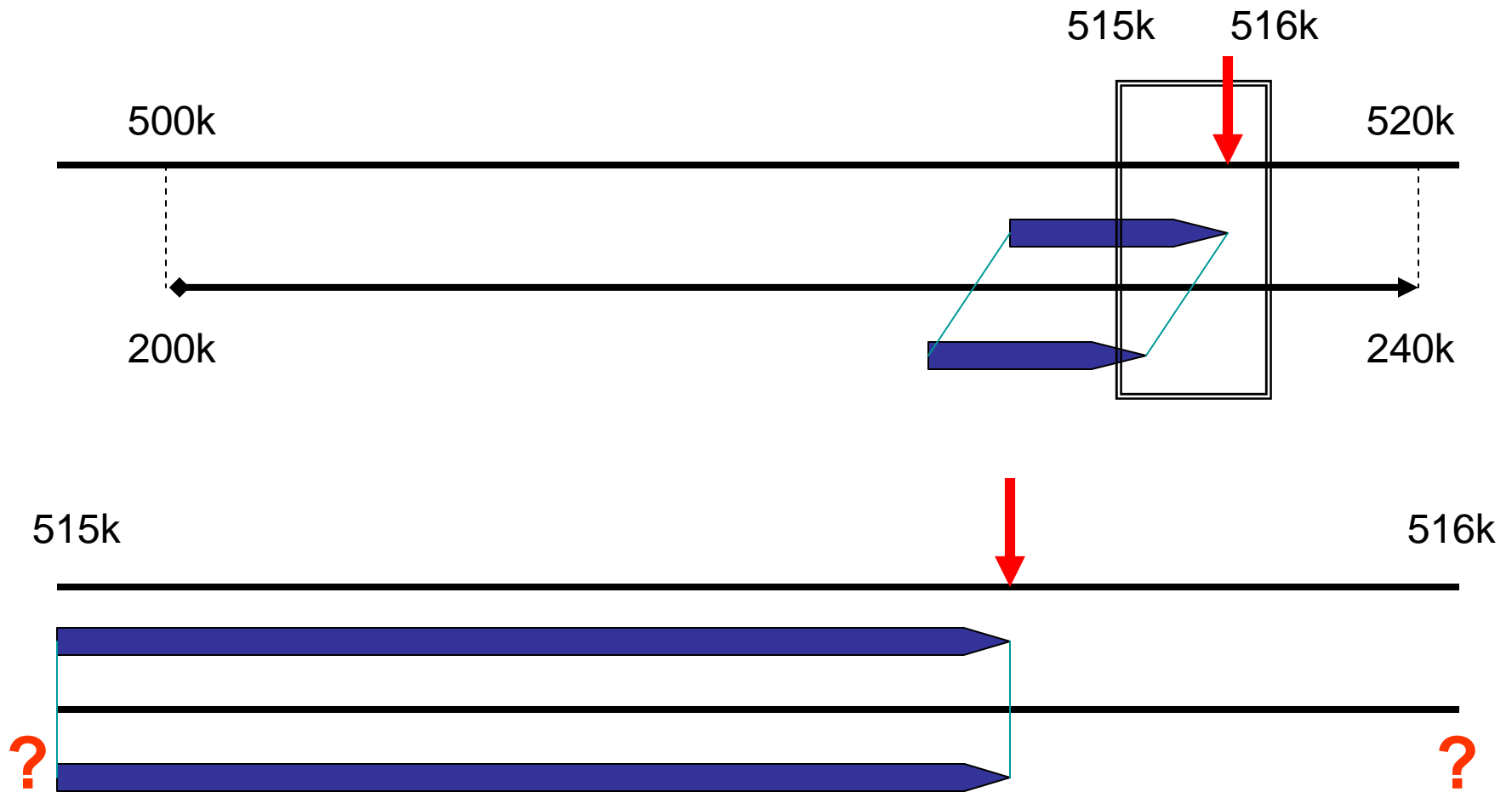
With a deletion in  
chromosome





Scale the syntenic gene 1:1 so that it shows up...

But what about the syntenic span?



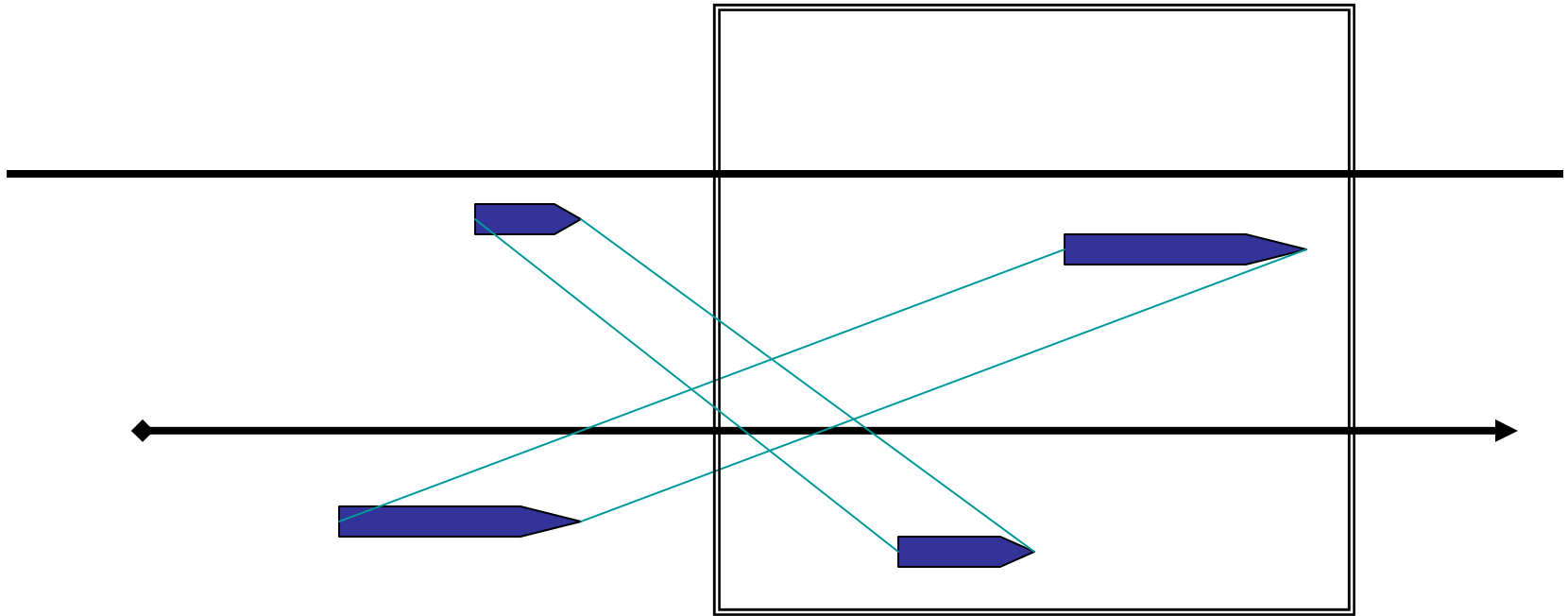
Only two choices:

1. 1:1
2. Use the full map

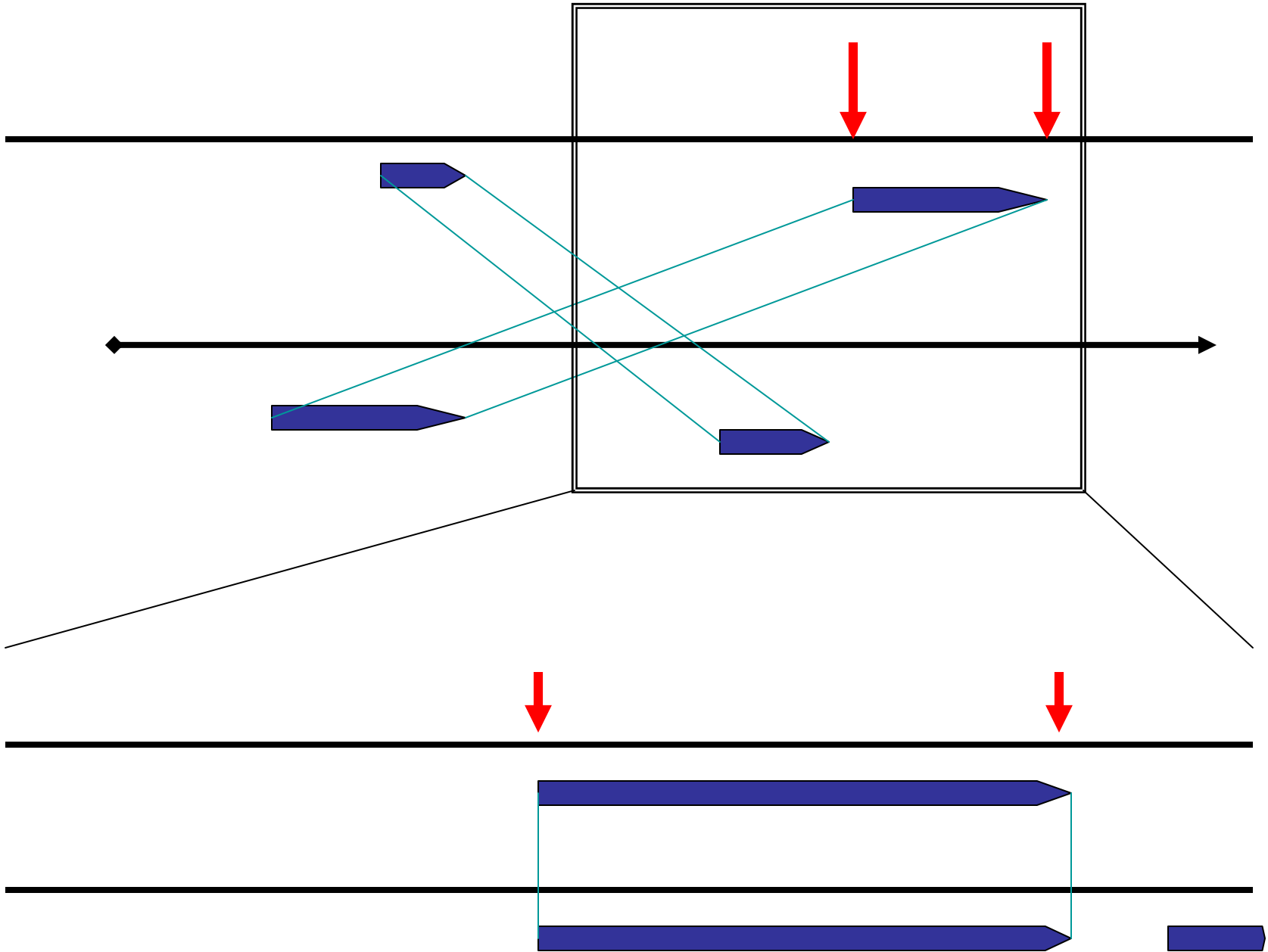
Each with artifacts:

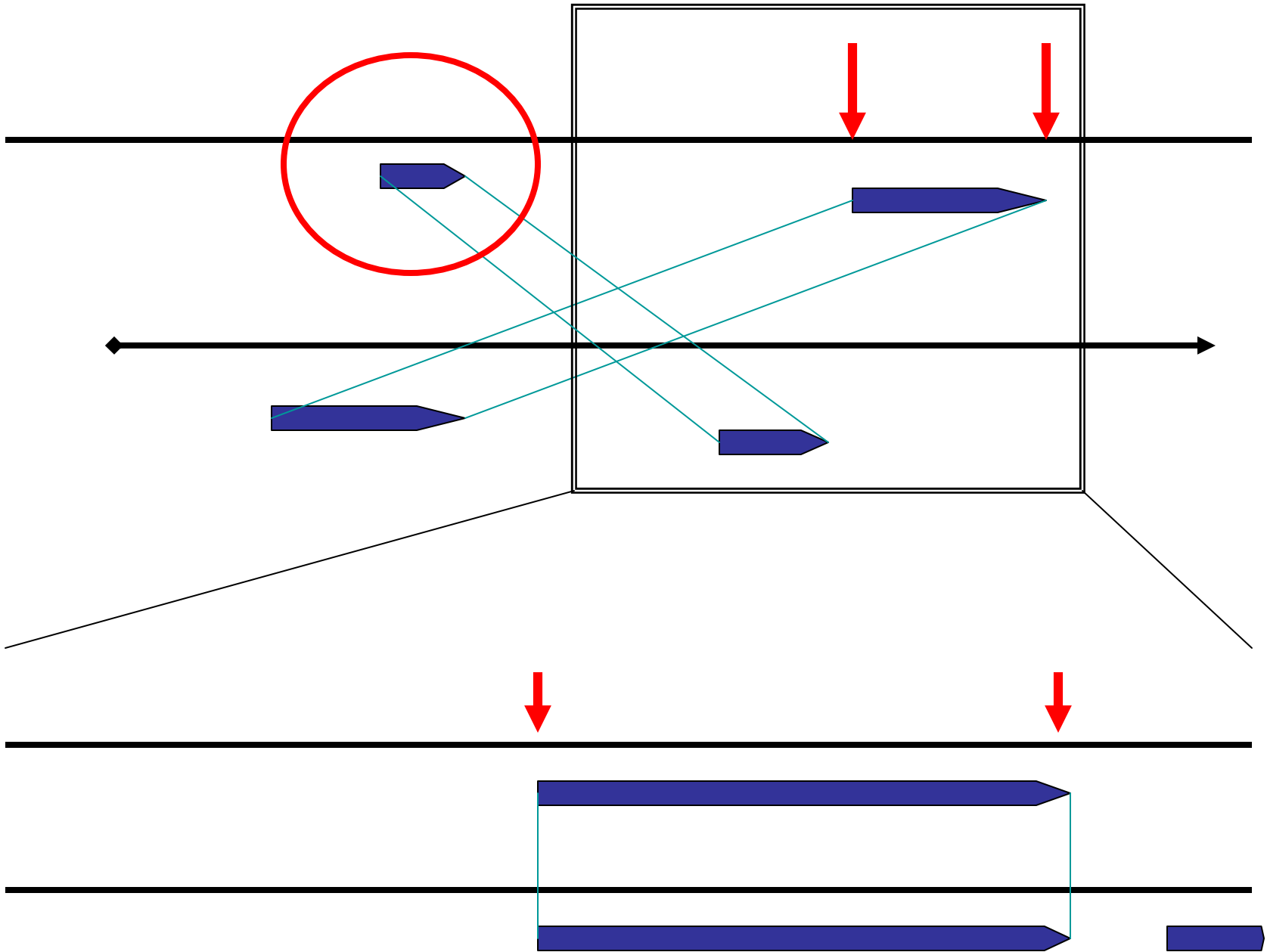
- might not be in display
- doesn't agree w/ genes





Another artifact. This one may be unavoidable.





# Synteny schema

## Synteny

	A_NA_SEQUENCE_ID	B_NA_SEQUENCE_ID	A_START	A_END	B_START	B_END	IS_REVERSED
1	233094	122440	1179350	1204720	2	20203	1
2	122440	233094	2	20203	1179350	1204720	1
3	233094	123222	1204725	1208262	3	1638	1
4	123222	233094	3	1638	1204725	1208262	1
5	233094	122400	1208268	1210404	4	1723	1
6	122400	233094	4	1723	1208268	1210404	1
7	233094	123102	1210414	1229444	4	8690	1
8	123102	233094	4	8690	1210414	1229444	1
9	233094	122565	1229461	1238994	1	3413	0
10	122565	233094	1	3413	1229461	1238994	0

## SyntenyAnchor

	SYNTENY_ANCHOR_ID	SYNTENY_ID	SYNTENIC_LOC	PREV_REF_LOC	REF_LOC	NEXT_REF_LOC
1	185577	110438	745884	563	860	999999999
2	185578	110438	749364	1	13	563
3	185579	110708	280827	660	1970	2066
4	185580	110708	282302	1	660	1970
5	185581	110708	288266	-999999999	1	660
6	185582	110708	279536	1970	2066	999999999
7	185583	110446	769760	42	698	757
8	185584	110446	771040	1	42	698
9	185585	110446	770474	-999999999	1	42
10	185586	110446	765424	698	757	999999999

# Acknowledgements

- Synteny viewer
  - Haiming Wang (UGA)
  - Aaron Mackey (GSK)
  - Brian Brunk (UPenn)
  - Many others on the ApiDB team
- OrthoMCL
  - Feng Chen (UPenn)
- Mercator
  - Colin Dewey (Wisc)
- Vision
  - David Roos (UPenn)
  - Chris Stoeckert (Upenn)
  - Jessie Kissinger (UGA)