Self Study: Yeast Genome Comparison

SESSION 4

MARTIN KRZYWINSKI
Genome Sciences Centre
BC Cancer Agency
Vancouver, Canada

EMBO PRACTICAL COURSE:
BIOINFORMATICS GENOME ANALYSES
Centre for Research & Technology - Hellas, Thessalonica, Greece
June 5–17, 2017

SESSION SETUP

Use what you have learned and create an image using data from previous day.

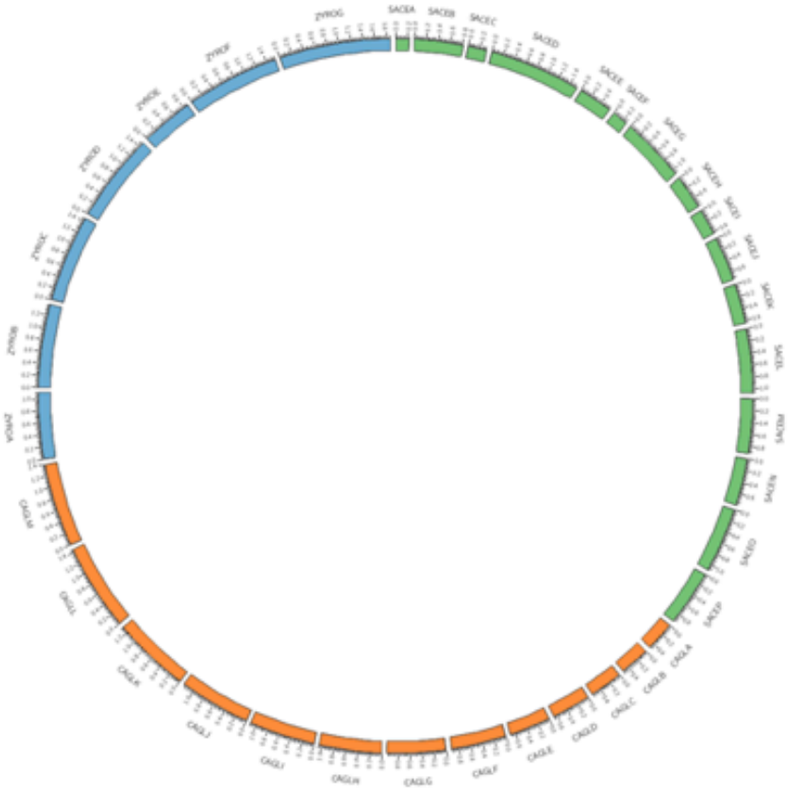
Input data is available in `session/4/data`

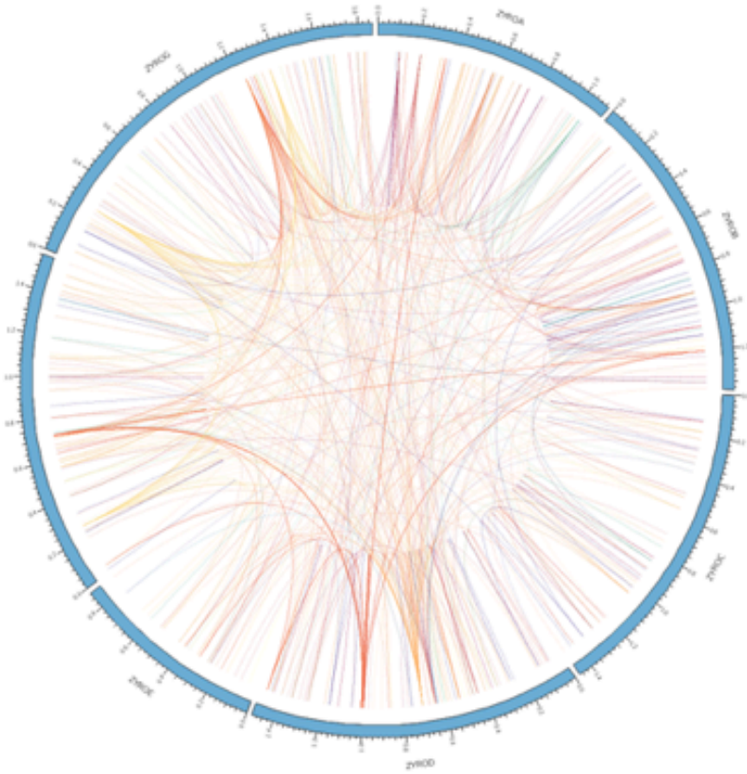
Each lesson starts you off with a template configuration `4/*/etc/circos.conf`

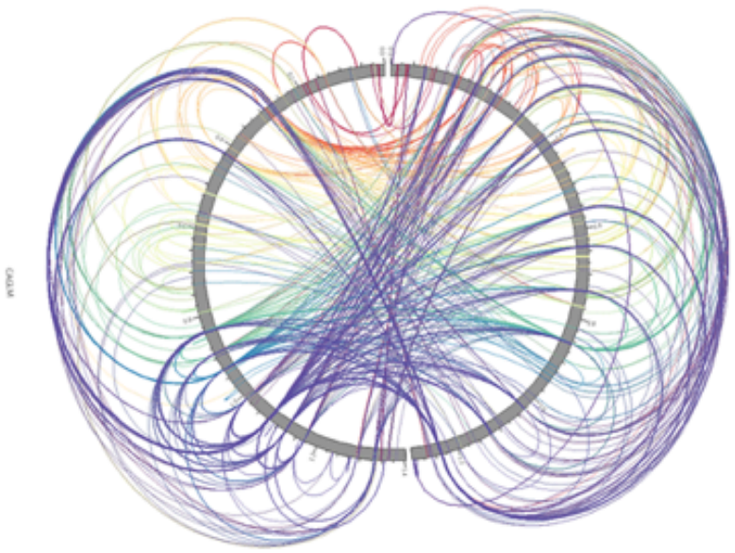
Follow the detailed handout (`handouts/session-4.pdf`) for this session to create the full configuration file. The instructions are also included in the template.

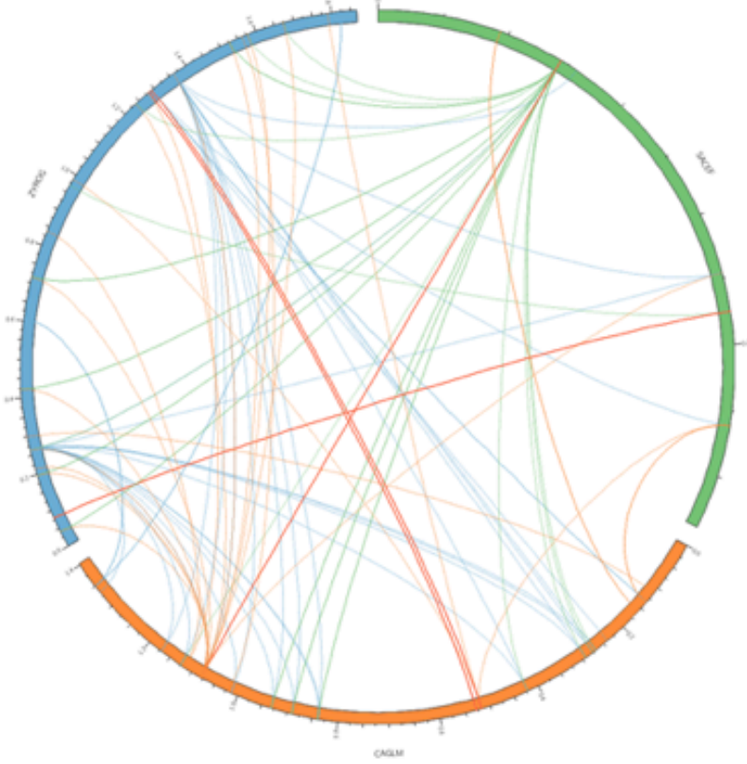
Answers are provided in *`session/4/*.solution/`*. Try your best before referring to them!

SESSION IMAGES







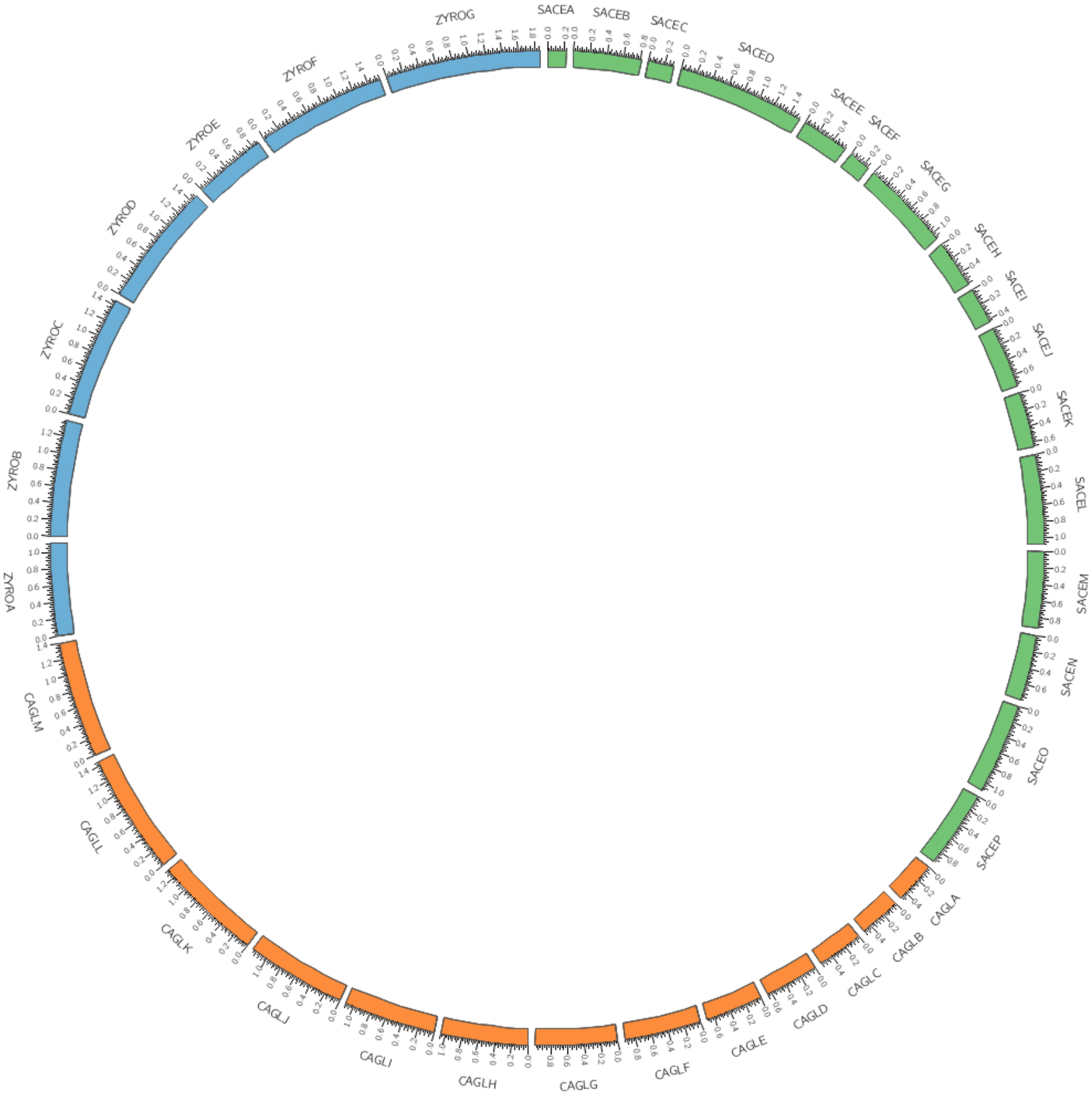


.

Yeast species comparison—drawing ideograms

LESSON 1

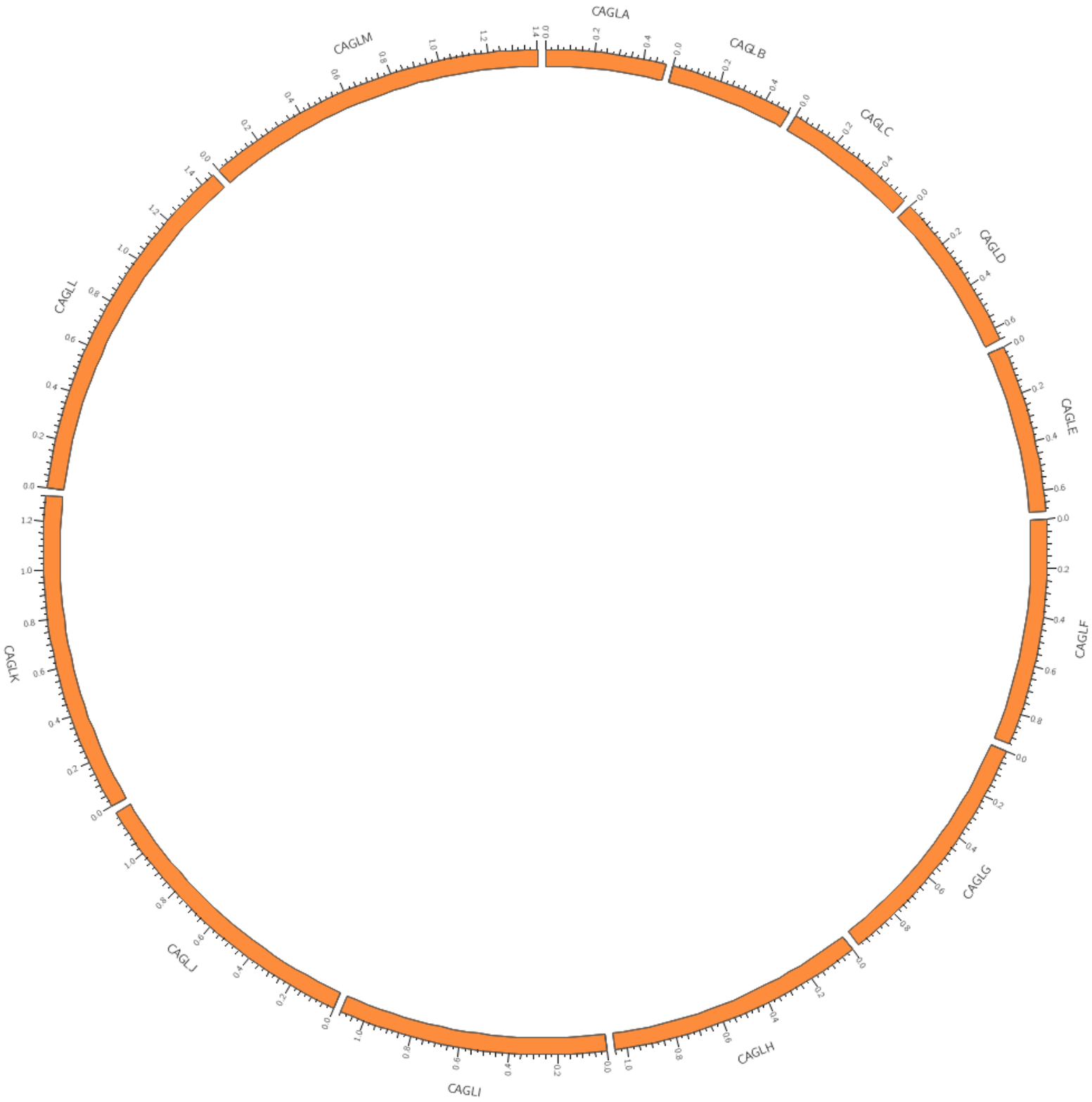
IDEOGRAM LAYOUT



Generate the image shown here showing all three genomes: SACE (green) CAGL (orange) and ZYRO (blue).

.

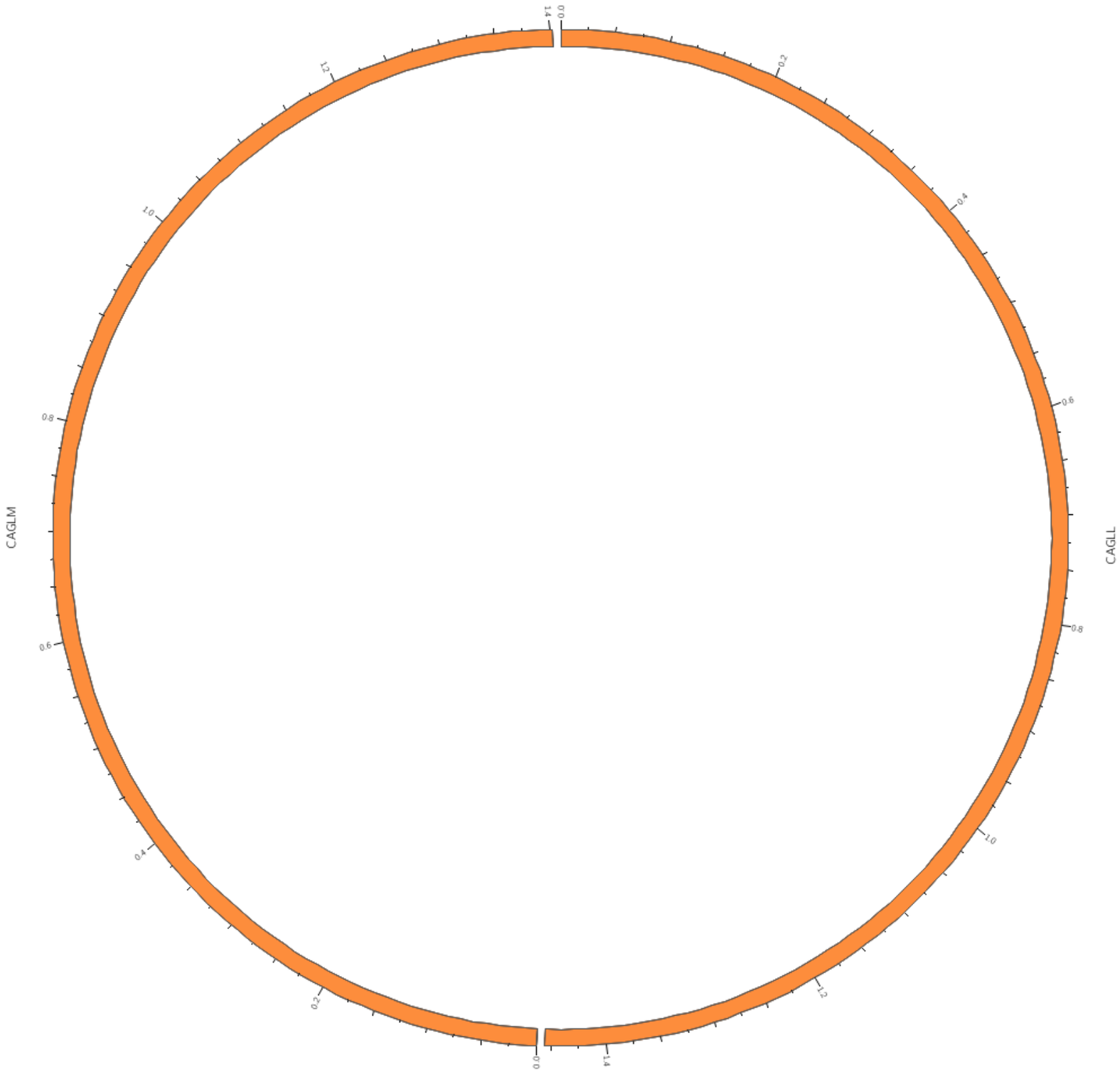
IDEOGRAM LAYOUT



Generate a version that shows only CAGL genome.

.

IDEOGRAM LAYOUT



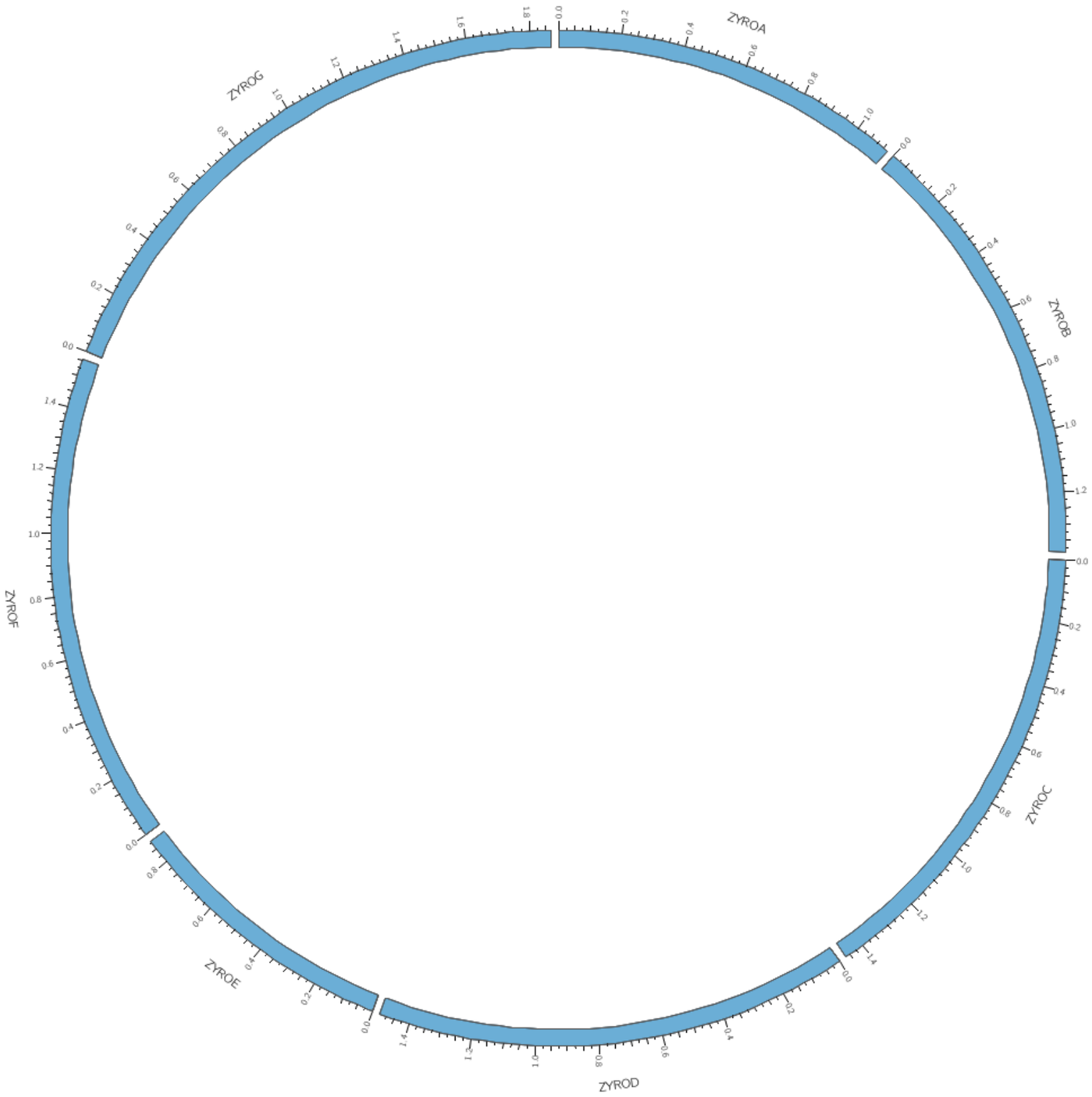
Generate a version that only shows `cagl-l` and `cagl-m` chromosomes, each occupying 1/2 of the image.

.

Yeast duplication—interior links

LESSON 2

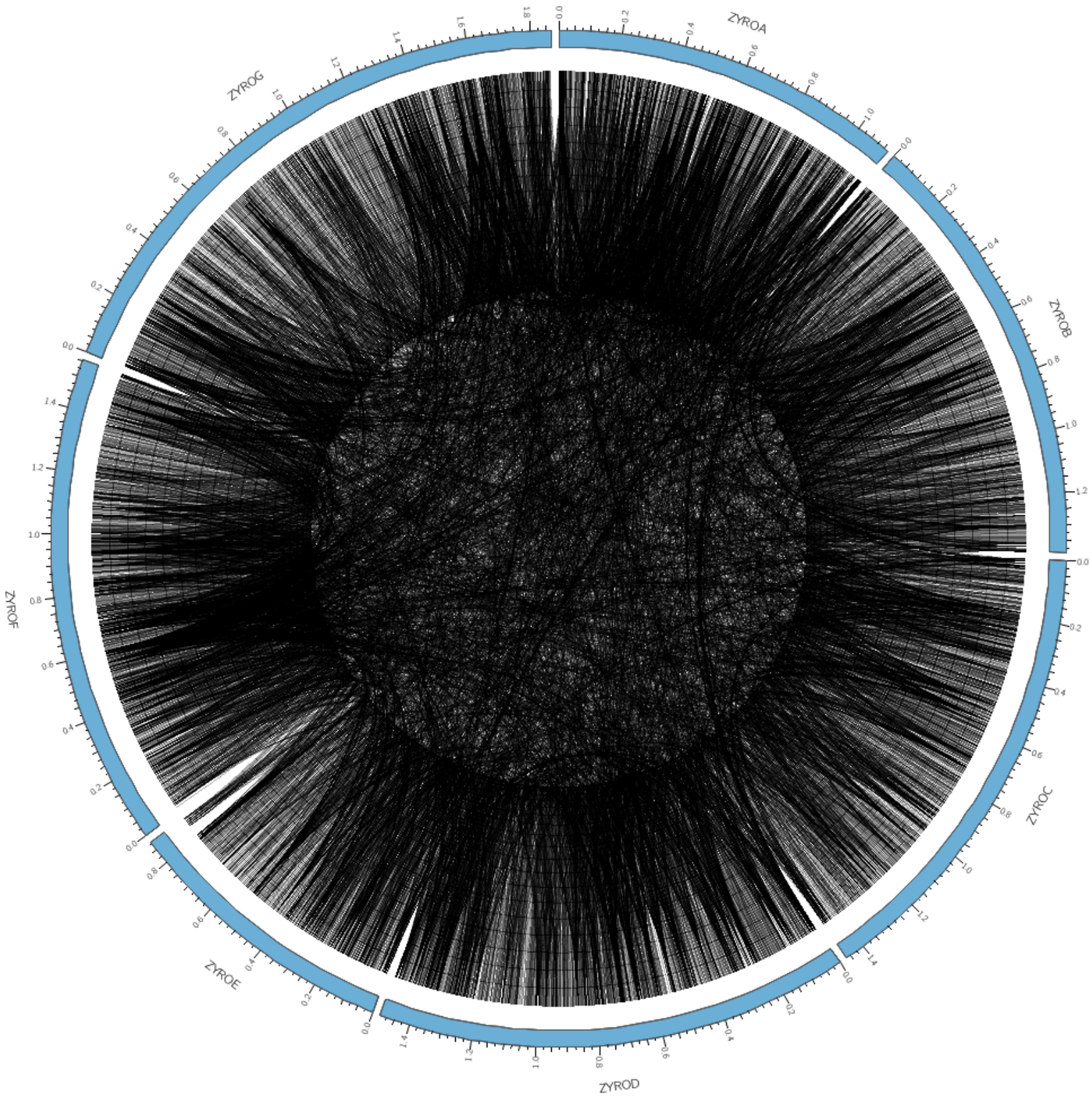
GENOME DUPLICATIONS



Draw the ZYRO genome with blue ideograms.

.

GENOME DUPLICATIONS



Draw links from the file

`CIRCOS/DUPLICATION/link_zyro_zyro`

with thickness 1, black and transparency level 5.

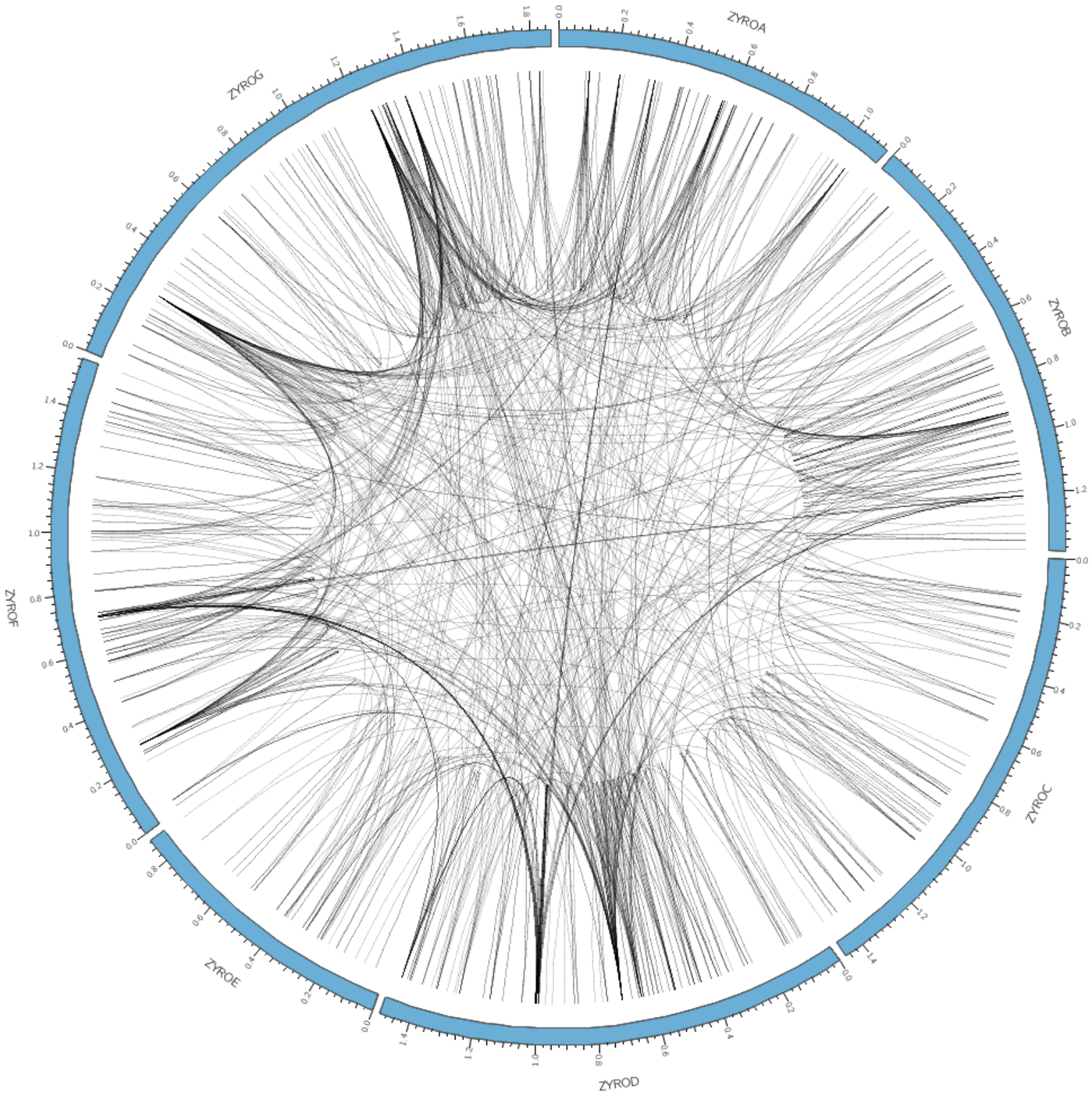
Use the `record_limit` parameter in the `<link>` block to load only a subset of links to speed up image generation during debugging.

e.g.

```
record_limit = 500
```

.

GENOME DUPLICATIONS

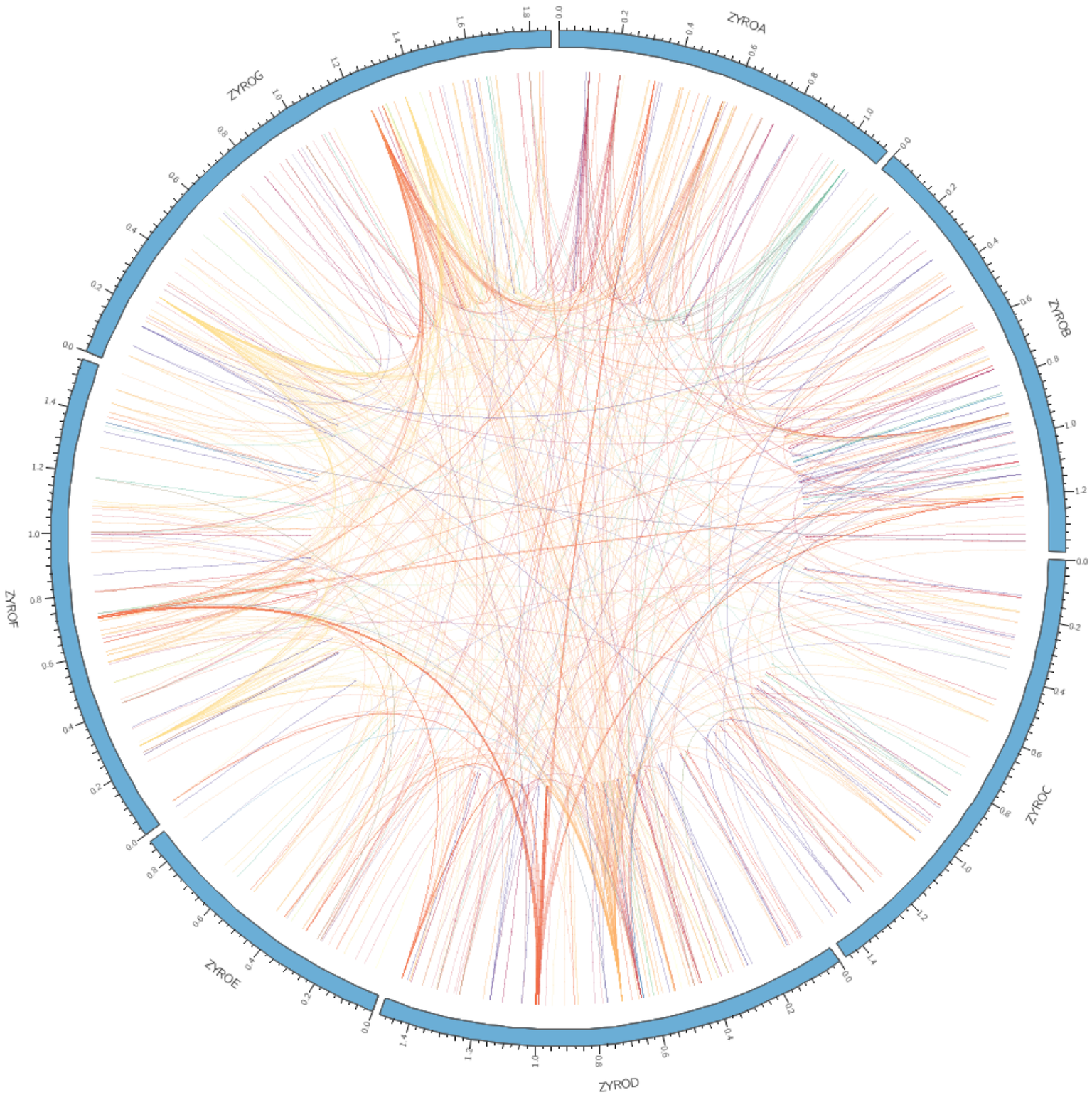


Add a rule that hides all links whose start coordinate is less than 4 kb in size.

You can access the start coordinate size using `var(size1)`.

.

GENOME DUPLICATIONS



Add another rule that changes the color of links based on their size.

Use the `spectral-11-div` palette and map size range 4-6 kb onto color index 1-11. Use `remap_int()` function for this.

```
remap_int(x,min,max,range_min,range_max)
```

Make the color transparent (e.g. level 5).

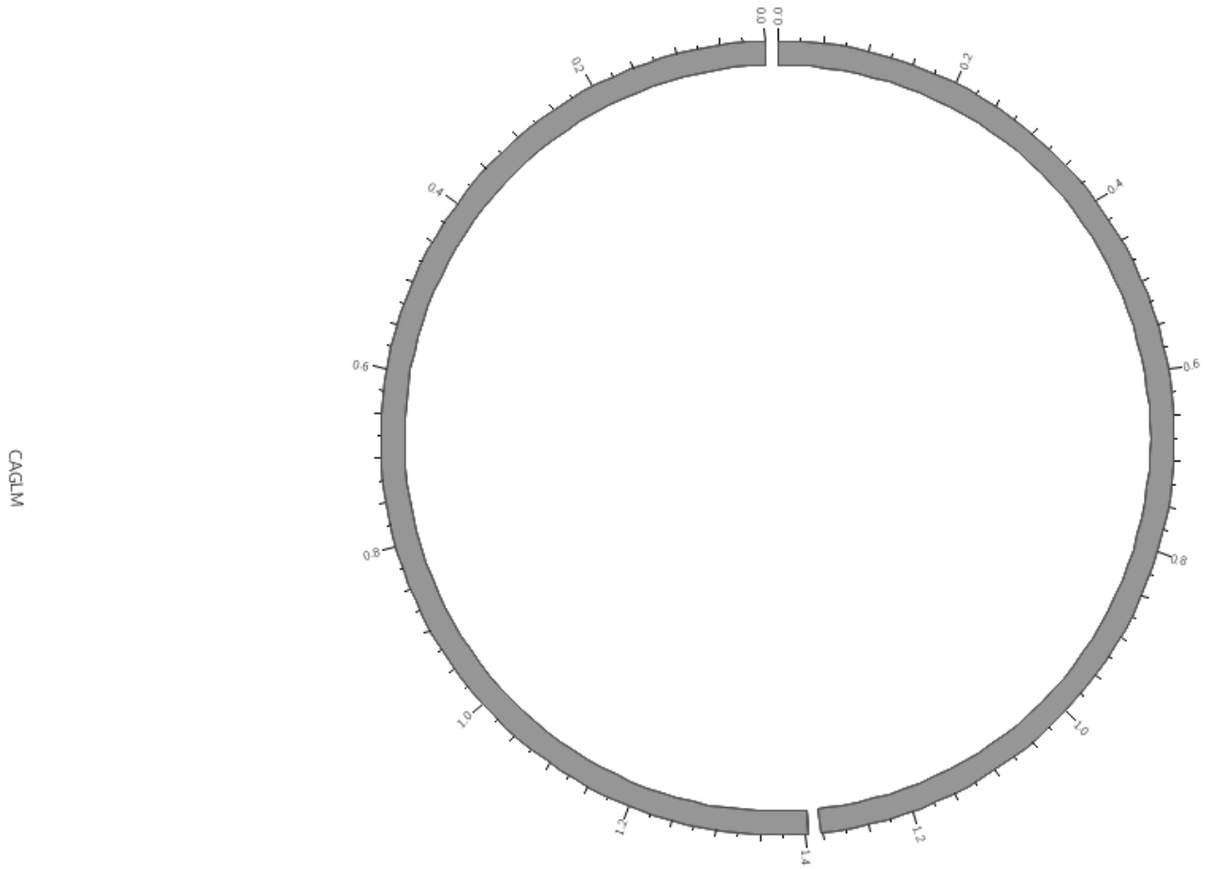
Set the `z` parameter in the rule so that larger links are drawn on top.

.

Yeast duplication—exterior links

LESSON 3

FOCUS ON DUPLICATIONS



Draw an image of `cagl-k` and `cagl-m` ideograms, each occupying 1/2 of the image.

Make ideograms grey.

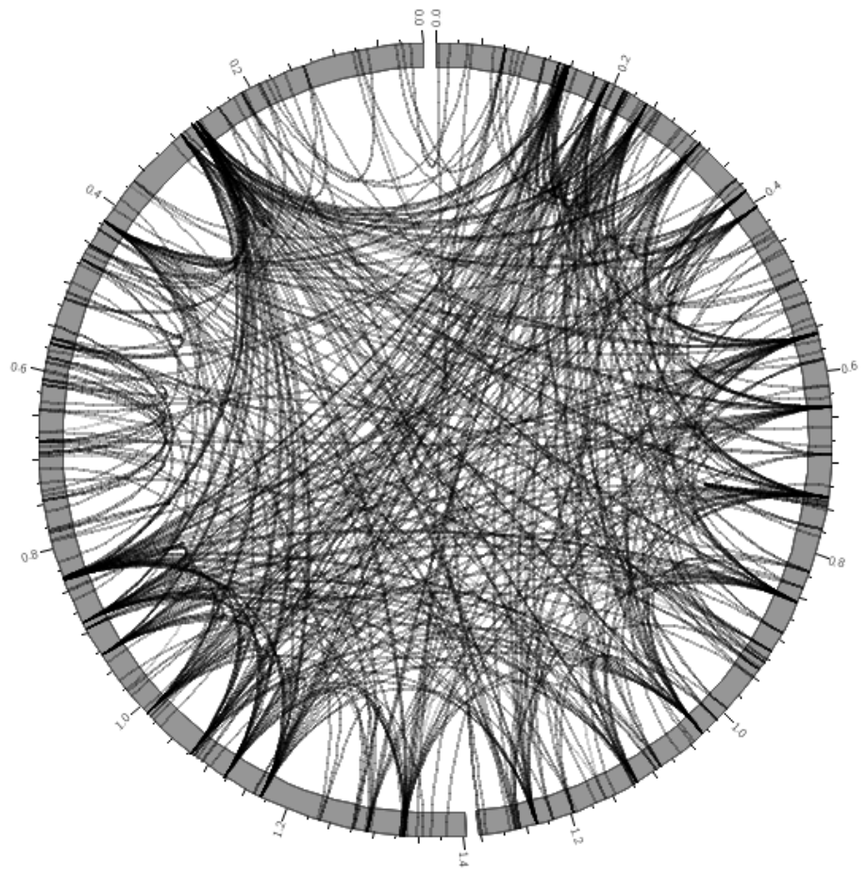
Make the ideogram radius `0.5r`.

Make the ideogram label radius `1.9r`.

Reverse orientation of `cagl-m`

.

FOCUS ON DUPLICATIONS



Draw duplications from

`CIRCOS/DUPLICATION/link_cagl_cagl`

as links of thickness 2 and black with transparency level 5.

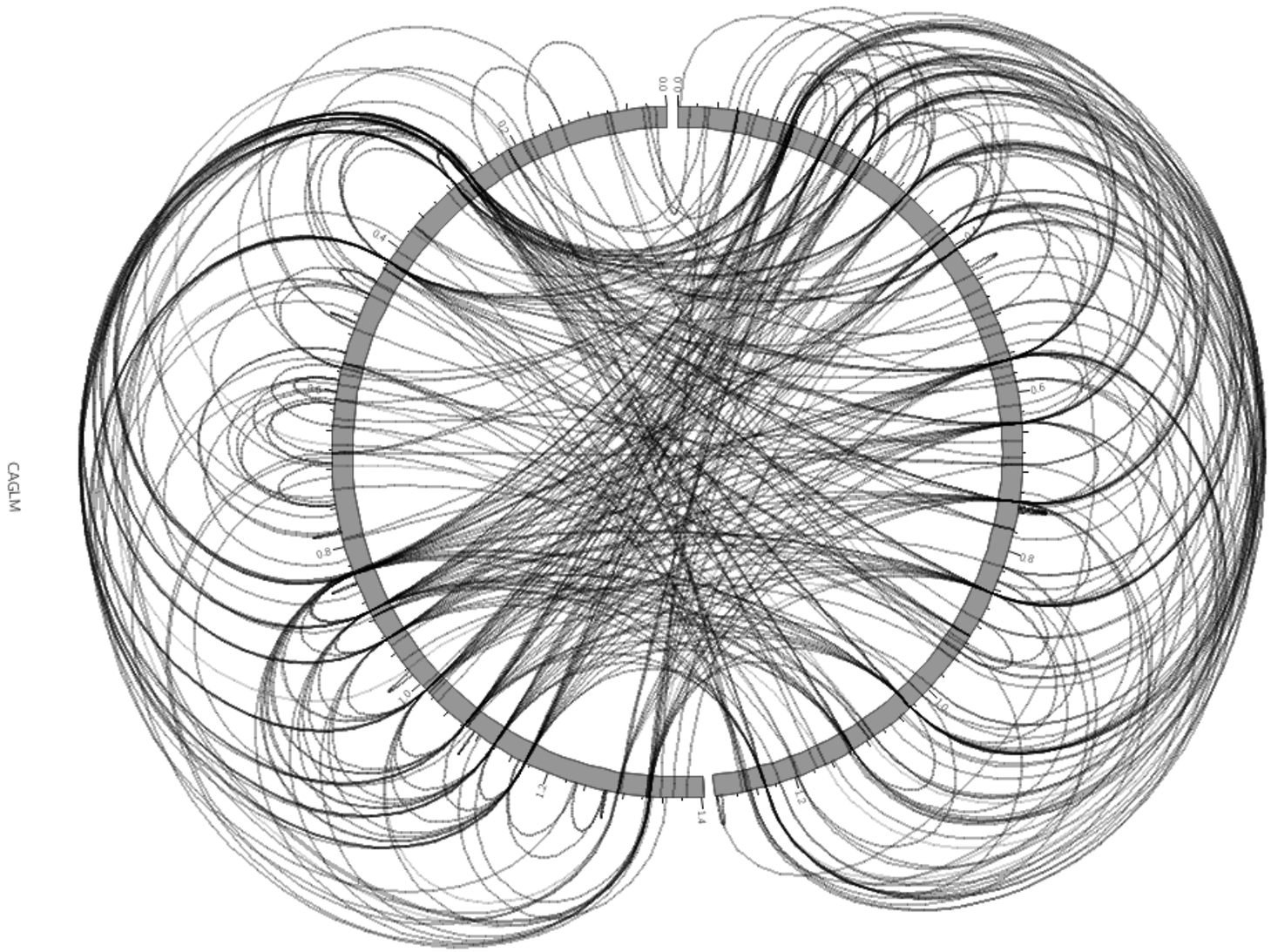
Set `radius` to `1r`.

Set `bezier_radius_purity` to `0.50`

Set `crest` to `0.5`.

Experiment with the last two parameters. What do they do?

FOCUS ON DUPLICATIONS



Create a rule that changes the `bezier_radius` for intrachromosomal links. Check this status in the rule condition using

```
var(intrachr)
```

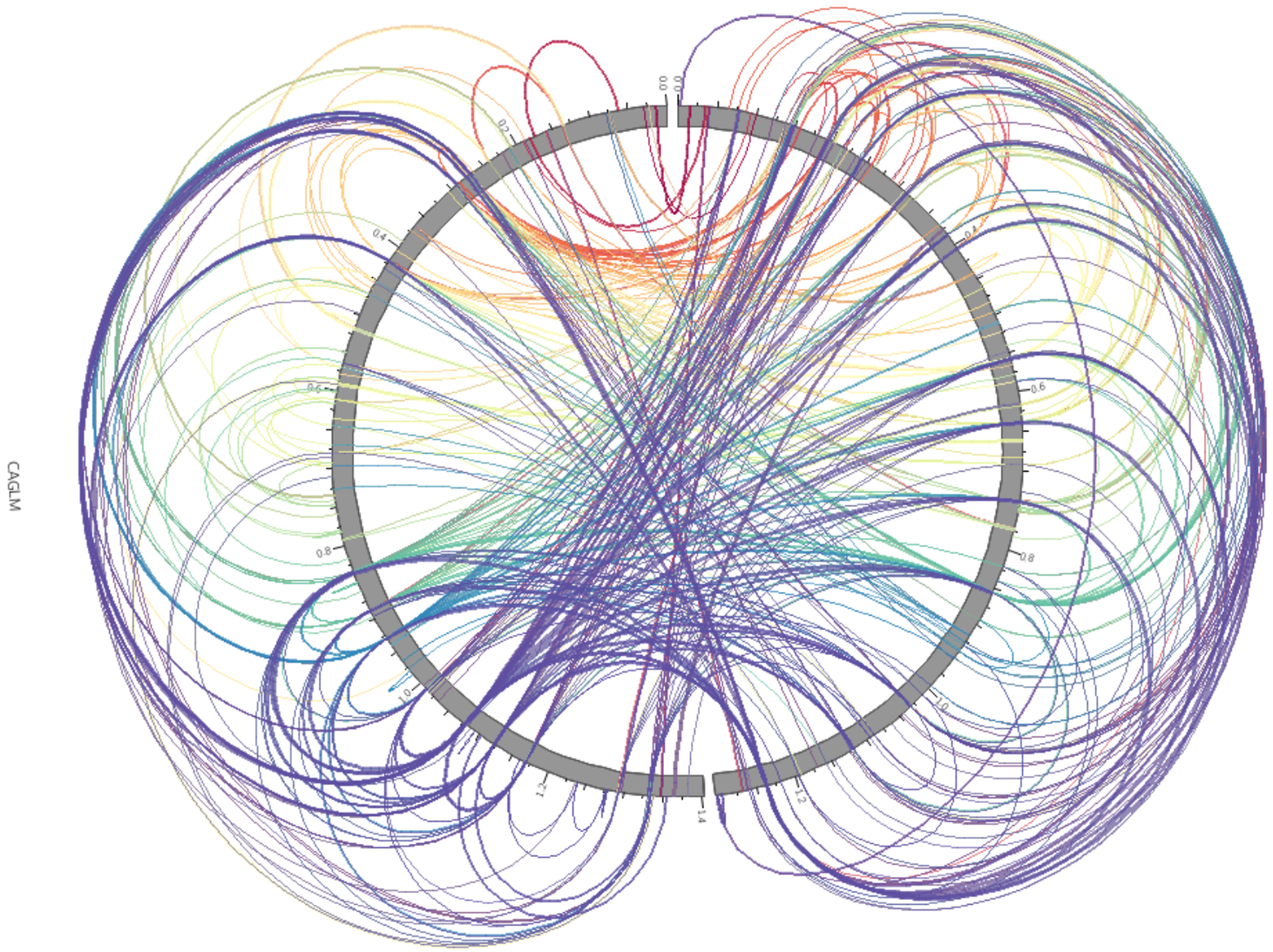
Remap the absolute difference between `start2` and `start1` (min=0, max=1e6) onto the range (1.25,6). Use `remap()`.

```
remap(x,min,max,range_min,range_max)
```

To continue processing the next rule even when this rule matches, set

```
flow = continue
```

FOCUS ON DUPLICATIONS



Add another rule that changes the `color`, `thickness` and `z` parameters of the link.

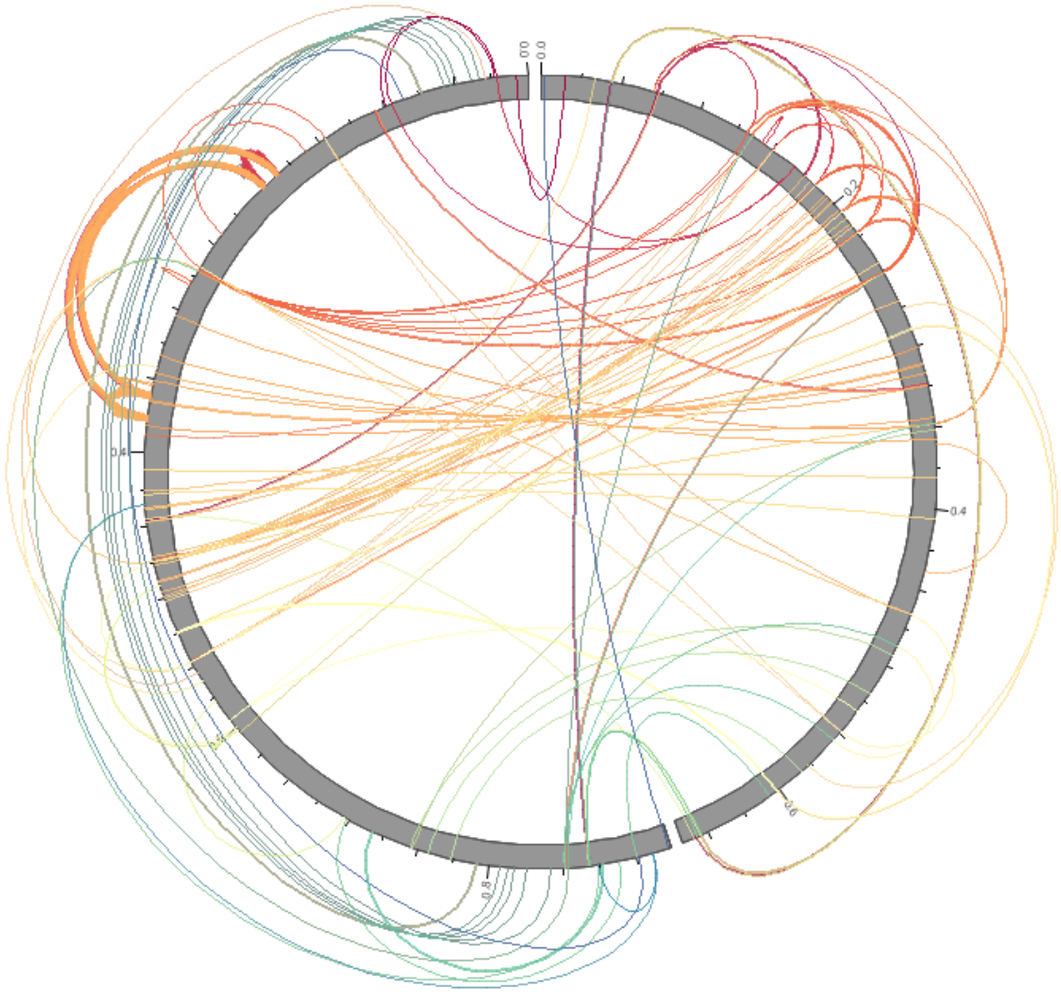
Assign `color` based on `start1` of link. Remap the start position `(0,1e6)` onto color index `(1,11)` and use `spectral-11-div` palette.

Assign `thickness` based on size of link start coordinate `(1000,5000)`. Map it onto thickness `(1,3)`.

Set the `z` parameter to be the start1 position.

.

FOCUS ON DUPLICATIONS



Define a parameter `genome` in the root of the configuration file.

You can access the value of this parameter using `conf(genome)` anywhere in the file.

Whenever you referred to `cagl` directly, use `conf(genome)`.

Change the parameter from `cagl` to `sace` to draw corresponding chromosomes in `sace`.

Now change the parameter to `zyro`. Did you see an error message? Try to figure out what it means. How would you fix the problem?

.

Yeast conservation

LESSON 4

GENOME CONSERVATION

Create a script in

`data/CIRCOS/CONSERVATION`

that extracts the 250 largest links from each `link_*` file (use the size of the start coordinate) and collects them into the file `links.top250.txt`.

Use bash for loop

```
for f in link_* ; do
...
done
```

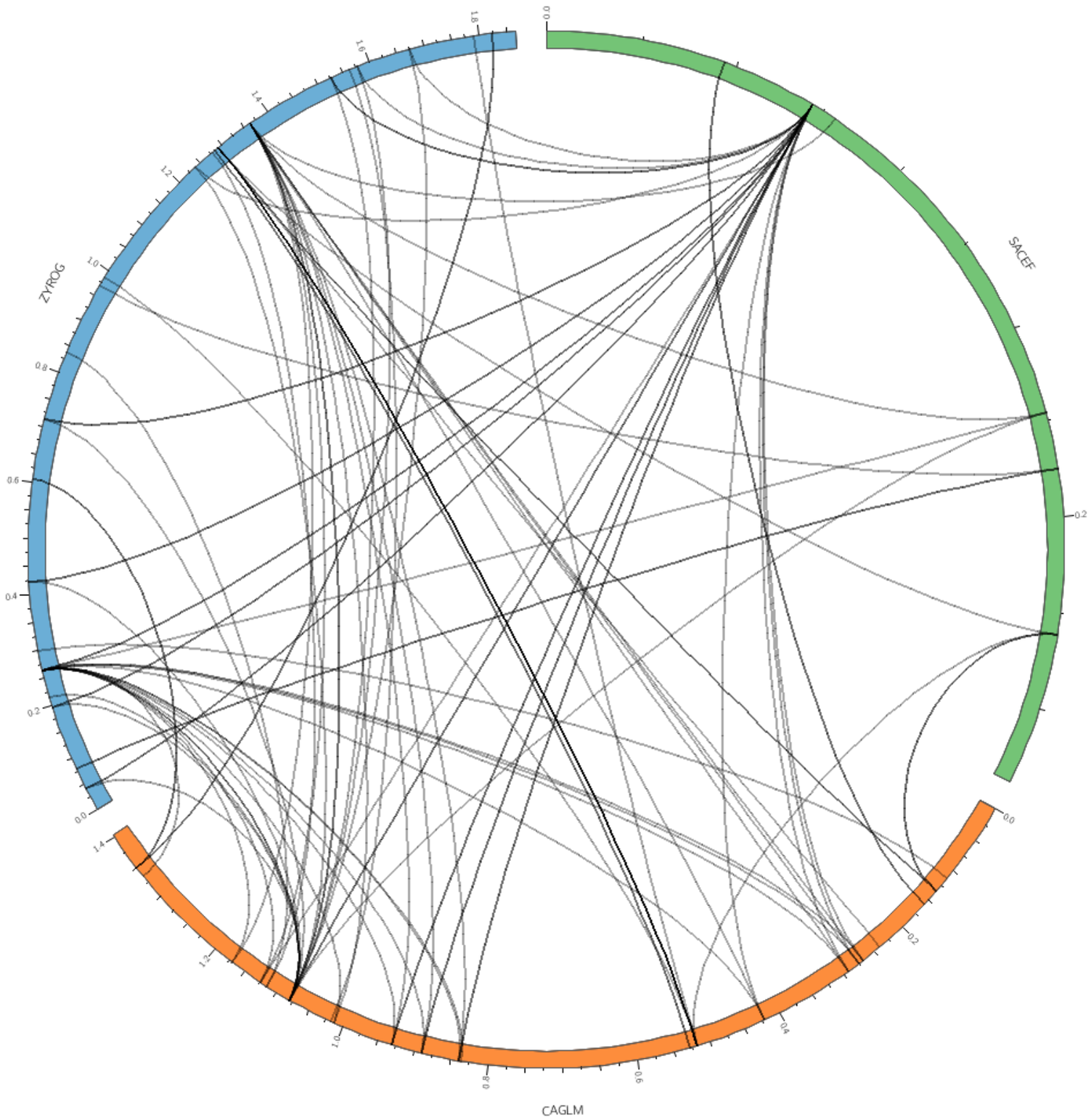
For the command, use `awk` to include the size of the difference to each line, then `sort` by this new field, then `head` to list only part of the file, then remove the field with `cut`.

The answer is in

`data/CIRCOS/CONSERVATION/topN`.

.

GENOME CONSERVATION

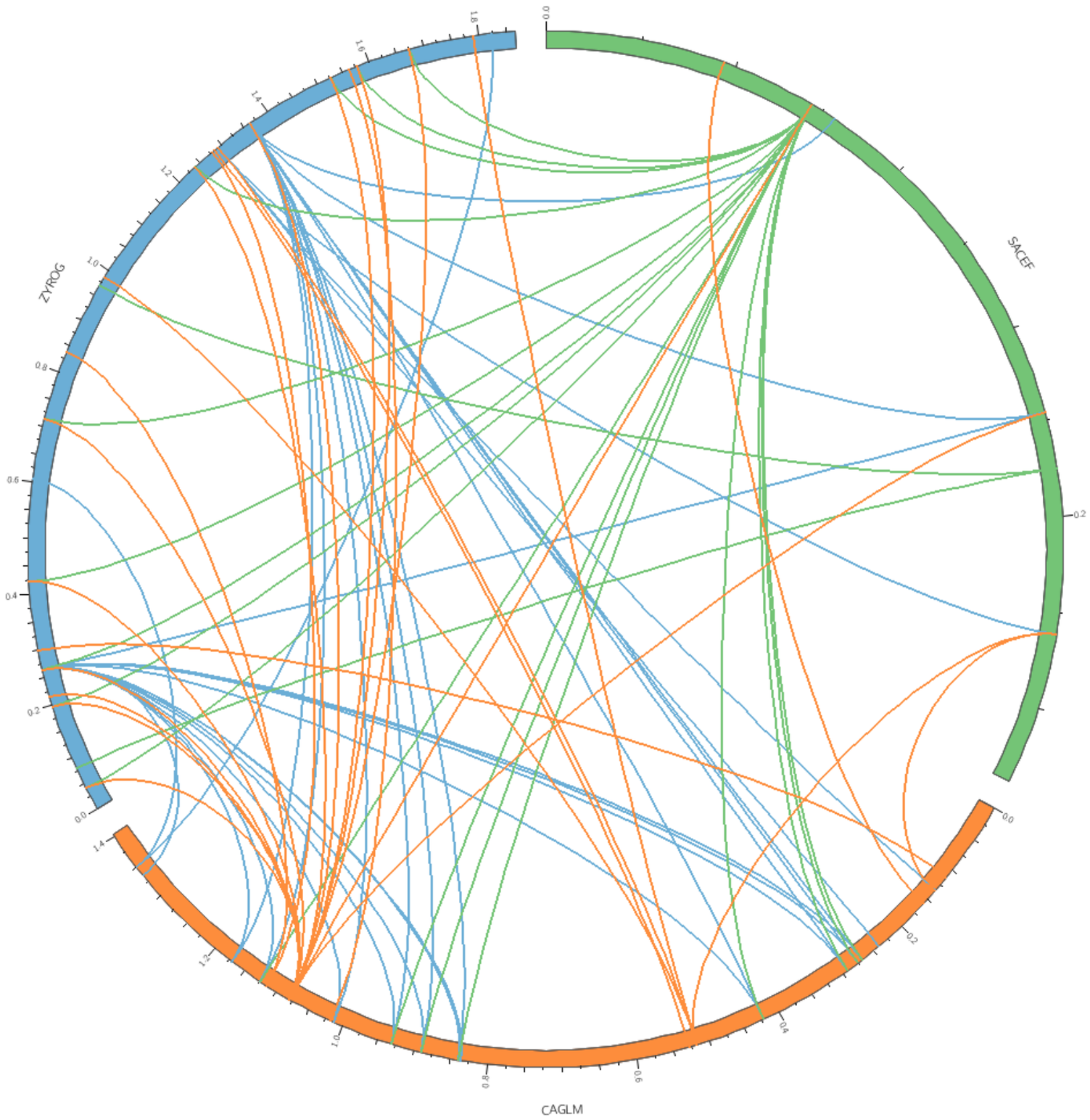


Draw `cagl-m`, `zyro-g` and `sace-f` ideograms. Make them each occupy 1/3 of the image.

Draw the links from the `links.top250.txt` file you created.

.

GENOME CONSERVATION



Set up rules that change the `color` of the link depending on what genome they originate from.

Use the `from(RX)` function in the rule condition to check whether the link starts on an ideogram that matches the regular expression `RX`.

Make links from CAGL orange, from SACE green and from ZYRO blue.

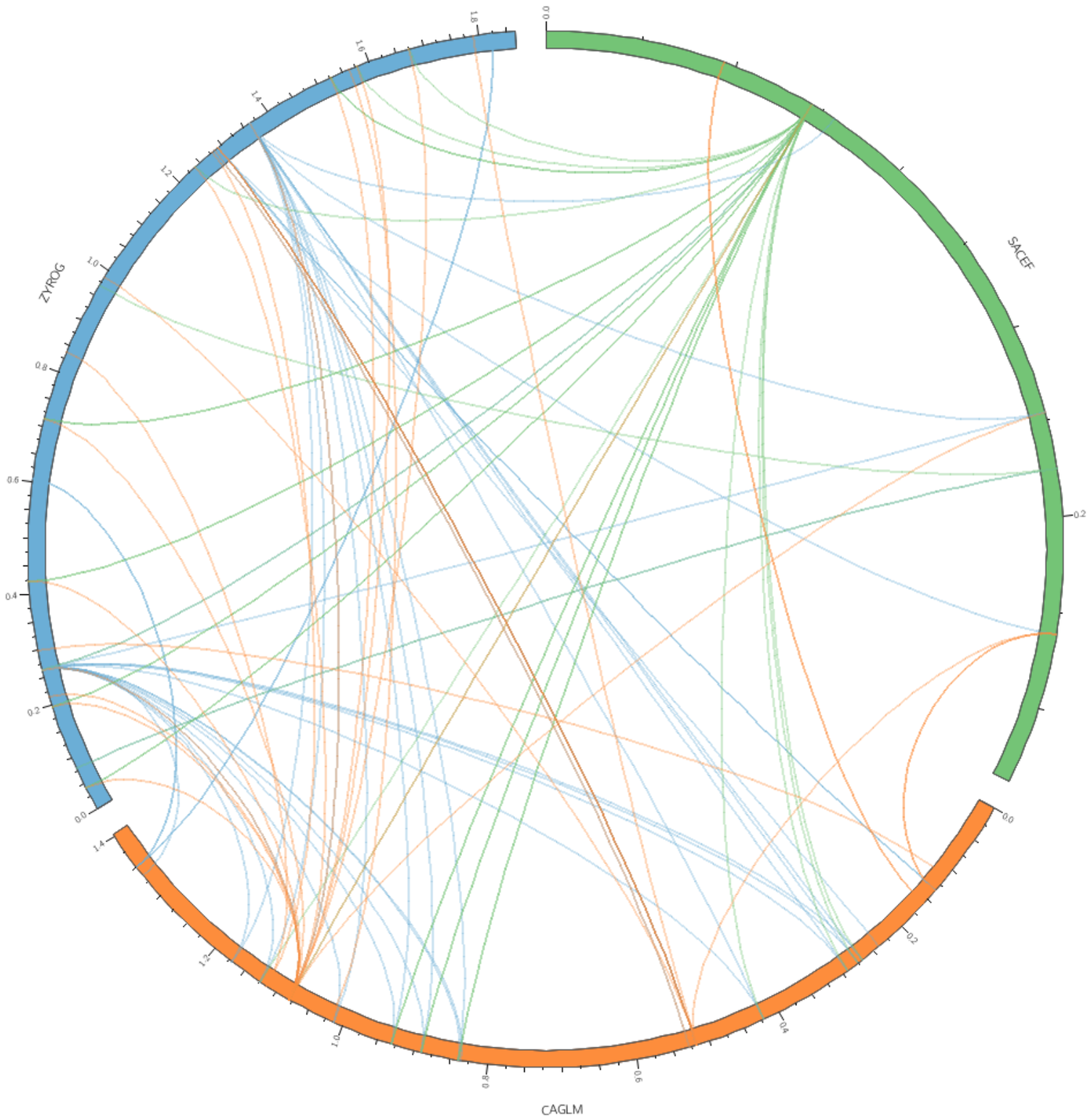
Set

`flow=continue`

globally for all rules. How does this help?

.

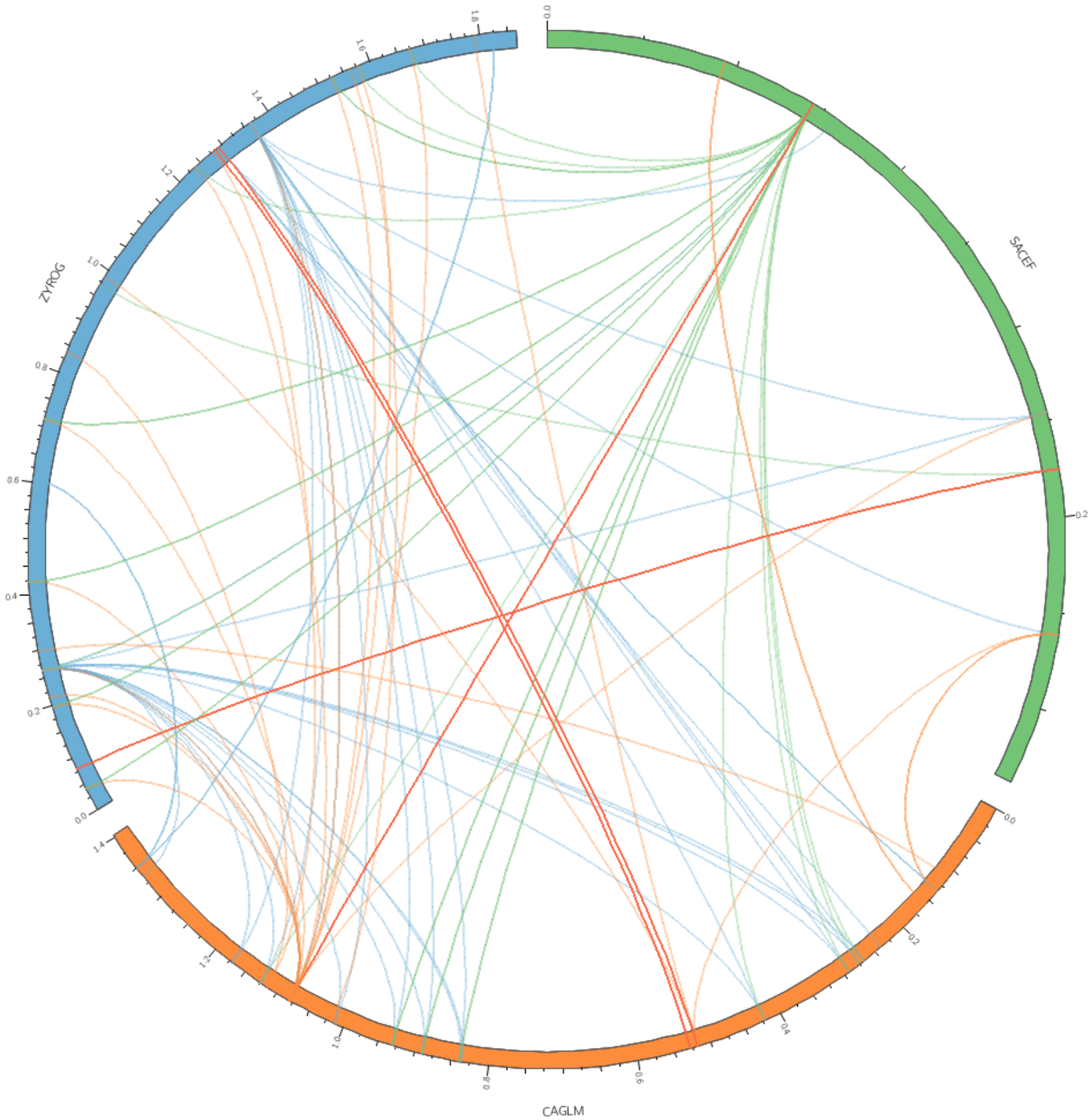
GENOME CONSERVATION



Add a rule that changes the color to a transparent version by adding _a4 to the end of the color name.

.

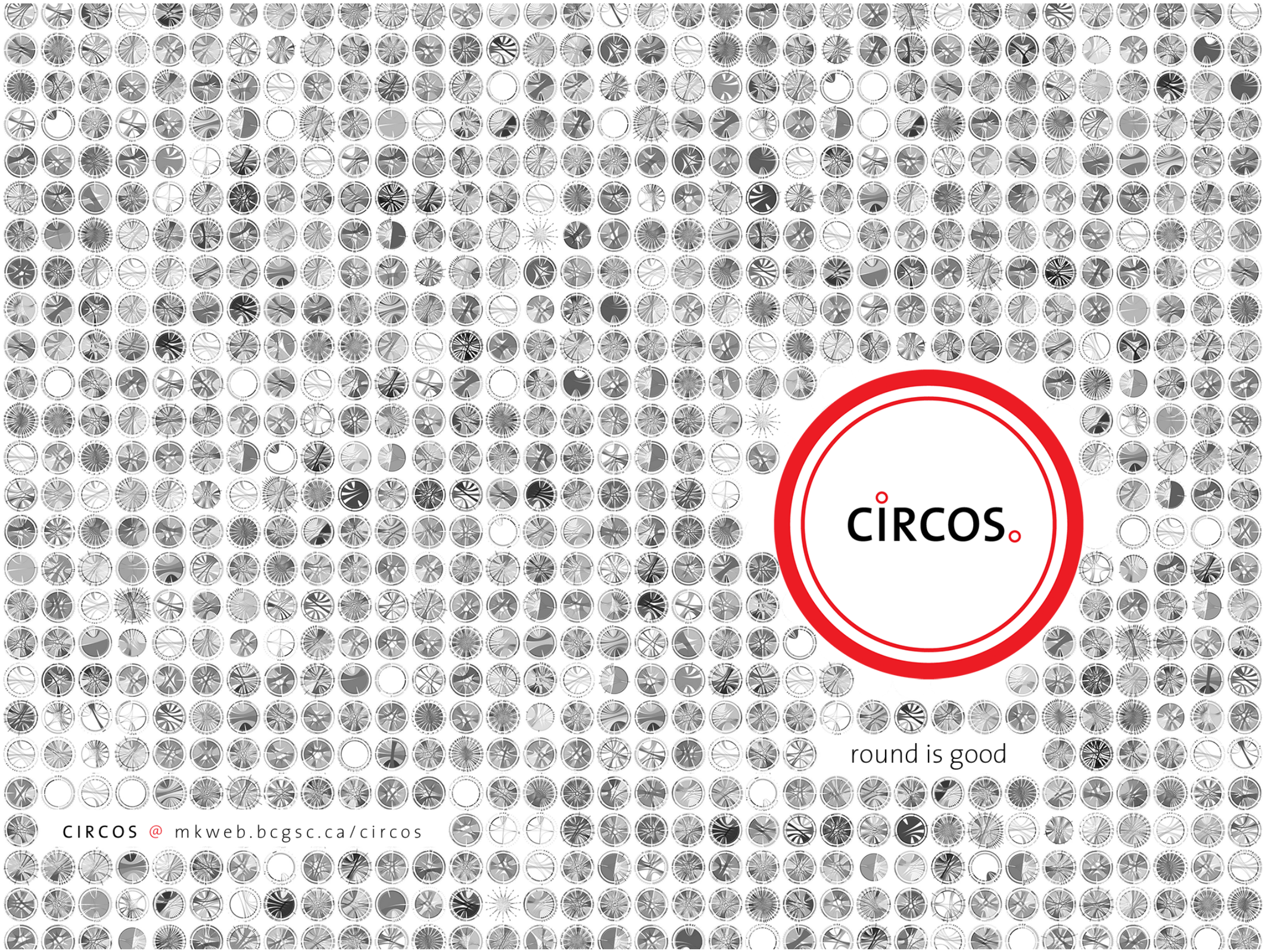
GENOME CONSERVATION



Add a rule that makes any links that have start and end coordinates larger than 5kb red.

Use `var(size1)` and `var(size2)` to access the link coordinate sizes.

.


CIRCOS.
round is good
CIRCOS @ mkweb.bcgsc.ca/circos

.