



Mapping out African
genomics research with 'sf'



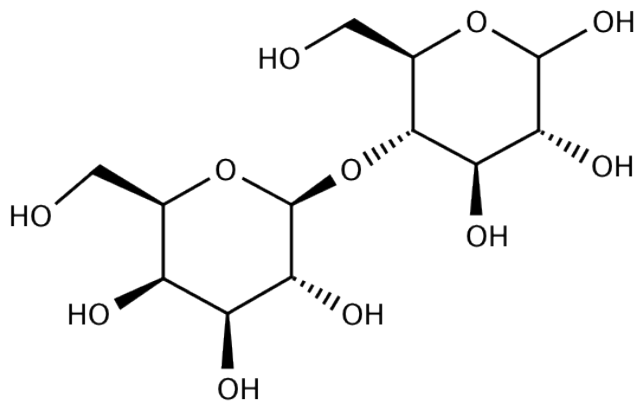
What is the Human Genome?



The collection of all of our genes,
the biological material
which shapes much of who we are

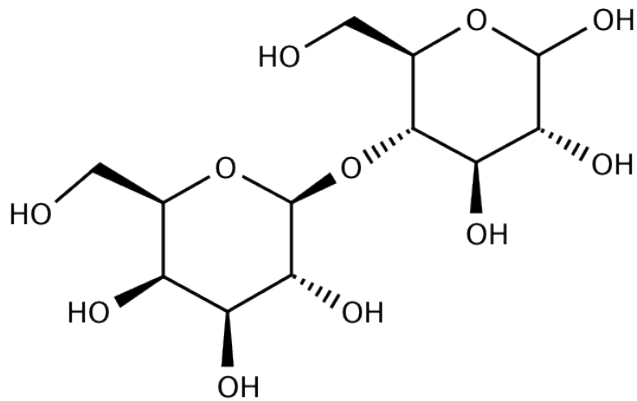
Our genes influence traits as diverse as:

- whether we can digest lactose

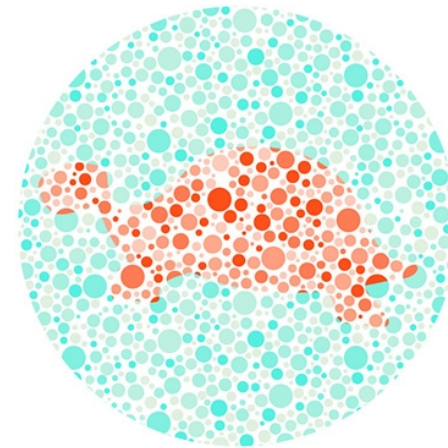


Our genes influence traits as diverse as:

- whether we can digest lactose

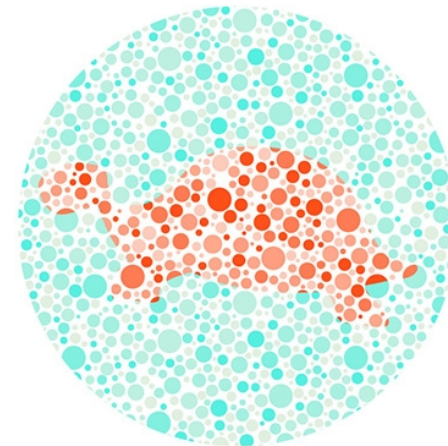
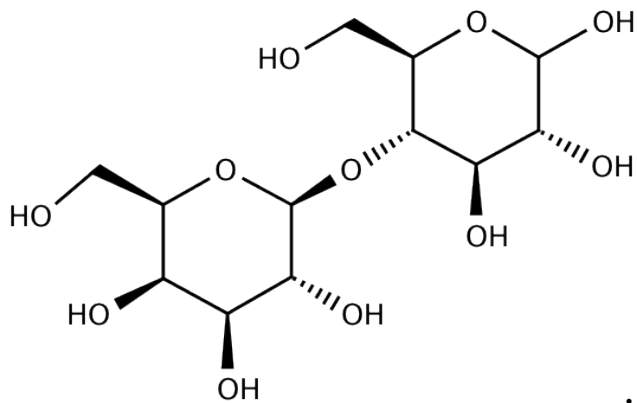


- our ability to distinguish pink from blue



Our genes influence traits as diverse as:

- whether we can digest lactose
- our ability to distinguish pink from blue



- our risk of developing heart disease



What makes African genome data so valuable?

- text

Recent studies exploring the human genome in African populations

Visualizing recent
progress using 'sf'

```
# load multiple packages in a single command using 'easypackages'
library(easypackages)
packages("sf", "dplyr", "ggplot2", "leaflet", "purrr")
```

```
# read in world borders data set (http://thematicmapping.org/downloads/world_borders.php)
map_data <- st_read(inDir, 'TM_WORLD_BORDERS-0.3.shp')
```

```
# subset african data with a bit of help from dplyr
map_data <- map_data %>%
  filter(REGION==2)
```

Simple feature collection with 6 features and 11 fields

geometry type: MULTIPOLYGON

dimension: XY

bbox: xmin: -8.667223 ymin: -18.01639 xmax: 31.30278 ymax: 37.09139

epsg (SRID): 4326

proj4string: +proj=longlat +datum=WGS84 +no_defs

	FIPS	ISO2	ISO3	UN	NAME	AREA	POP2005	REGION	SUBREGION	LON	LAT	geometry
1	AG	DZ	DZA	12	Algeria	238174	32854159	2	15	2.632	28.163	MULTIPOLYGON (((2.96361 36...
2	AO	AO	AGO	24	Angola	124670	16095214	2	17	17.544	-12.296	MULTIPOLYGON (((11.75083 -1...
3	BN	BJ	BEN	204	Benin	11062	8490301	2	11	2.469	10.541	MULTIPOLYGON (((2.484418 6...
4	CF	CG	COG	178	Congo	34150	3609851	2	17	15.986	-0.055	MULTIPOLYGON (((12.77905 -4...
5	CG	CD	COD	180	Democratic Republic of the Congo	226705	58740547	2	17	23.654	-2.876	MULTIPOLYGON (((12.95305 -5...
6	BY	BI	BDI	108	Burundi	2568	7858791	2	14	29.887	-3.356	MULTIPOLYGON (((29.2299 -3...


```
# read in sampling info
sampling_info <- (read.csv(paste0(inDir, "/sampling_info.csv")))
```

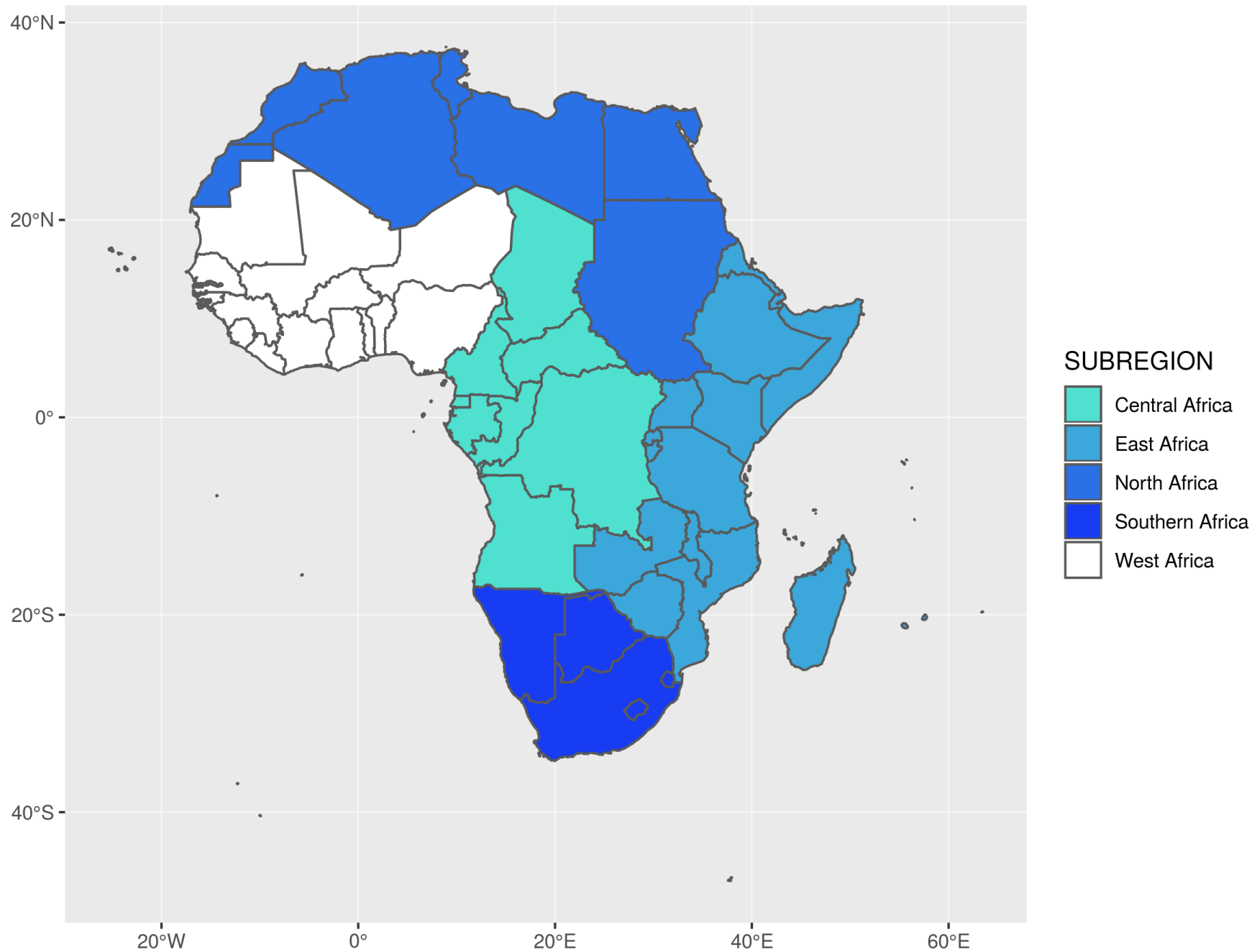
```
# add coordinates derived from map_data
sampling_data <- full_join(sampling_info, sampling_coordinates, by = 'NAME')
```

	NAME	MAG	LON	LAT
1	Uganda	100	32.3860	1.28000
2	South Africa	100	23.1210	-30.55800
3	Ethiopia	120	39.6160	8.62600
4	Nigeria	99	8.1050	9.59400
5	Gambia	113	-15.3860	13.45300
6	Kenya	101	37.8580	0.53000
7	Sierra Leone	85	-11.7920	8.56000
8	Nigeria	109	8.1050	9.59400
9	South Africa	8	23.1210	-30.55800
10	South Africa	8	23.1210	-30.55800
11	Burkina Faso	34	-1.7400	12.27800
12	Mali	50	-3.5240	17.35000
13	Nigeria	49	8.1050	9.59400
14	Ghana	26	-1.2070	7.96000
15	Benin	50	2.4690	10.54100
16	Botswana	48	23.8150	-22.18200
17	Zambia	41	26.3200	-14.61400
18	Cameroon	50	12.2770	5.13300
19	Other	400	102.3446	7.36965
20	Guinea	46	-10.9420	10.43900
21	Cote d'Ivoire	40	-5.5560	7.63200
22	Democratic Republic of the Congo	23	23.6540	-2.87600
23	Uganda	50	32.3860	1.28000

```
# quickly find a colours in between
colorRampPalette(colors=c("Blue", "Turquoise"))(5)
```

```
# plot data
ggplot(map_data) + geom_sf(aes(fill = SUBREGION)) +
```

[illegible]



```
p <- ggplot() +  
  geom_sf() + # add country polygons, shaded by SUBREGION  
  geom_label() + # add labels above or below country midpoints  
  scale_fill_manual() + # hand-select colours  
  scale_color_manual()
```

```
q <- p + # specify map variables  
  theme() +  
  scale_x_continuous() +  
  scale_y_continuous() +  
  guides() +
```

```
r <- q +  
  geom_point() # add sampling data
```

```
r %>% ggsave() # export figure
```

