

Myths and facts on migration in Europe and the global context

Nikola Sander

Vienna Institute of Demography



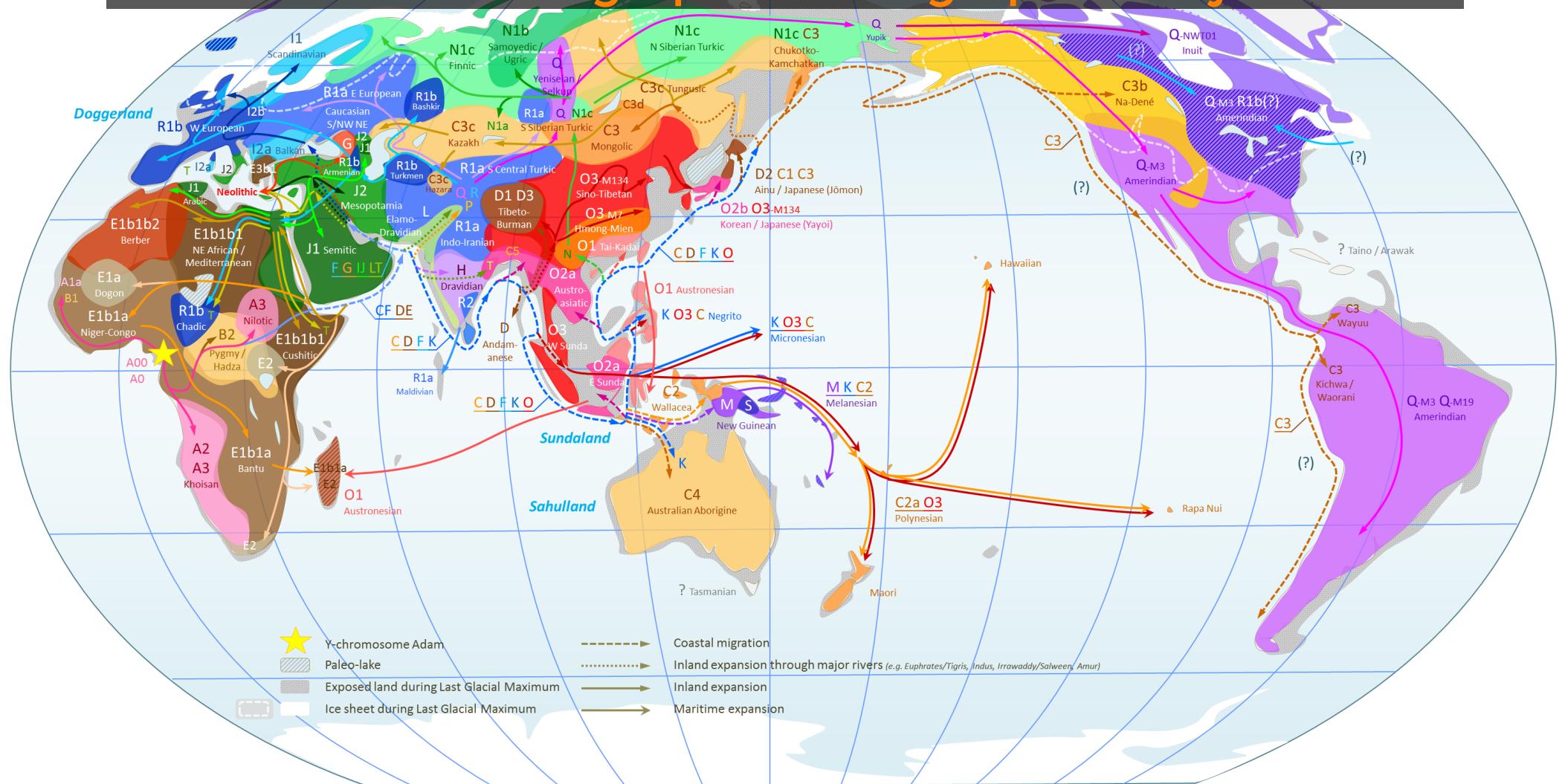
People have migrated for 60,000 years

World Map of Y-Chromosome Haplogroups Y-DNA migrants for 60

emigrated for 60 years, and the descendants of the original settlers now number over 100 million.

with Possible Migrations Routes

National Geographic Genographic Project

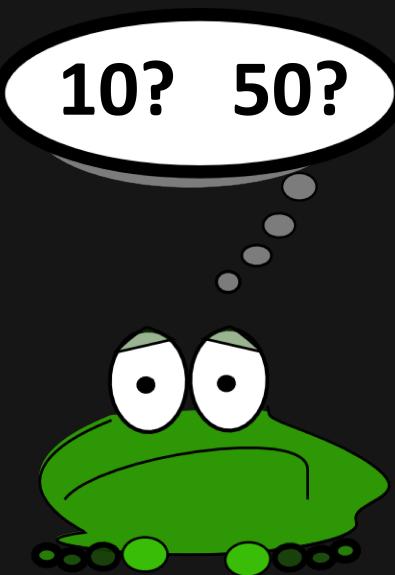


Why people migrate

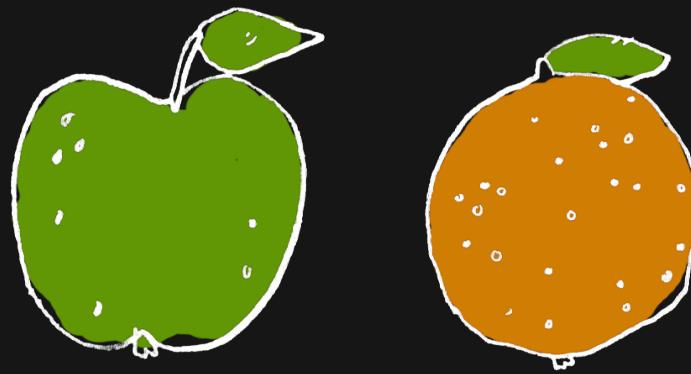
1. economic (eg higher wages)
2. family (eg marriage)
3. housing (eg larger flat)
4. education (eg attending university)
5. violent conflict

A rising tide of mass migration?

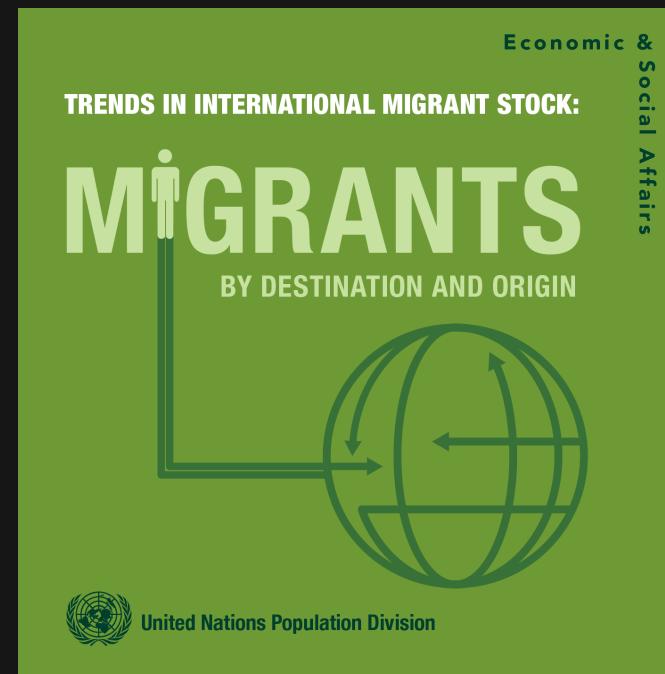
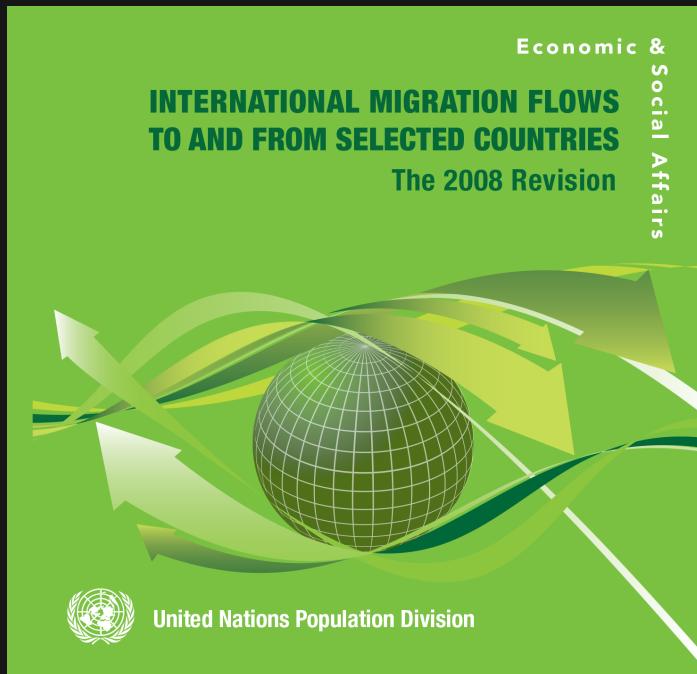
For every 1,000 people worldwide, how move over 5-year periods?



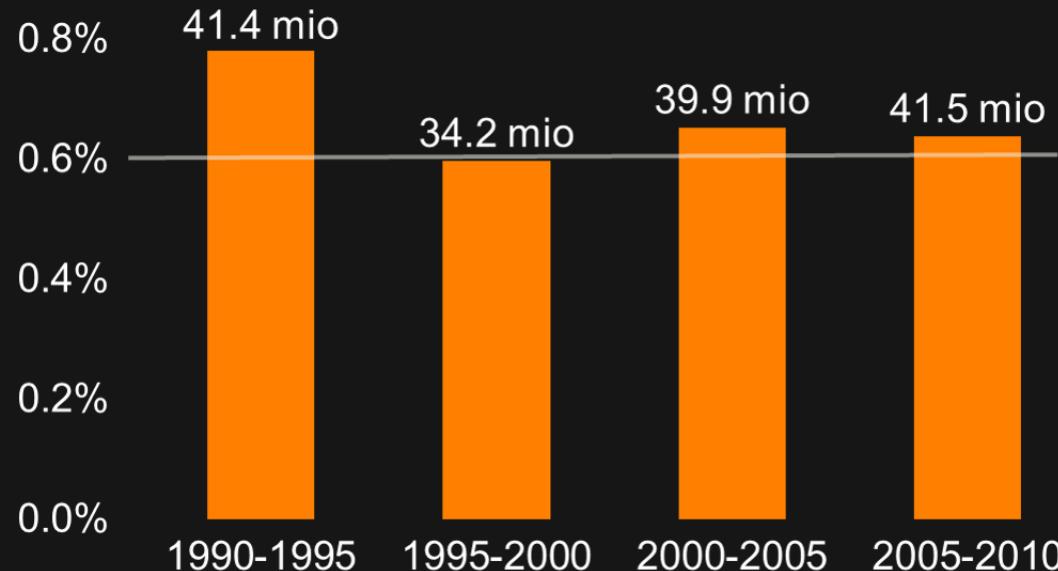
Migration can be defined in many different ways



Two types of migration data: flows & stocks



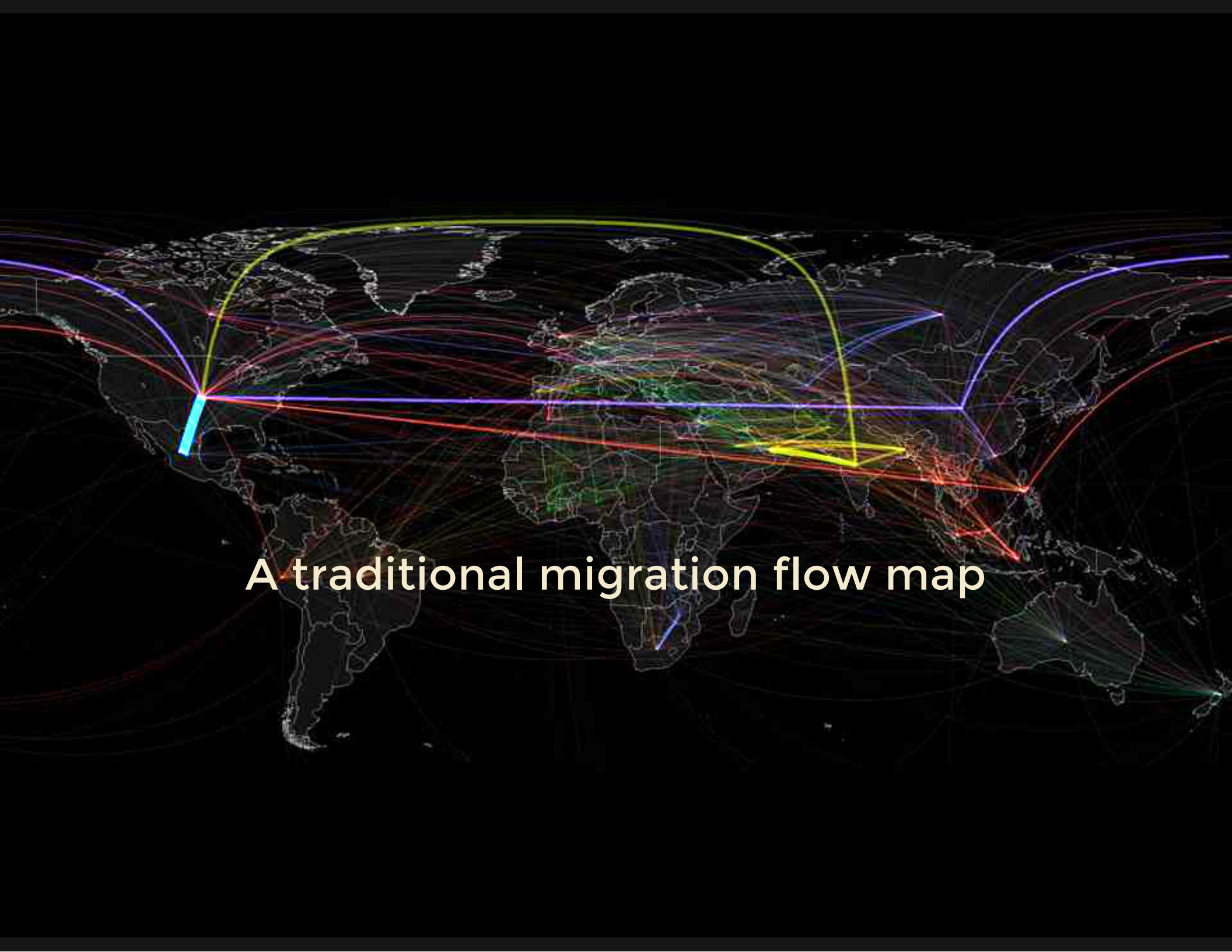
Our estimates: **0.6%** of the world's population move over 5-year periods



Abel & Sander (2014)
Quantifying Global International Migration Flows
Science, vol. 343: 1520-1522.

How to explore a matrix of flows between 196 countries?

How to explore a matrix of flows between 196 countries?

A world map with a dark background, showing numerous thin, multi-colored lines representing migration flows between countries. The lines are concentrated along coastlines and major routes, such as from South America to North America and Europe, and from Africa to Europe and the Middle East. Two specific paths are highlighted with thicker lines: one originating in South America and another in Africa, both ending in North America.

A traditional migration flow map

Circos: a source of inspiration

Canada's Michael Smith Genome Sciences Centre // British Columbia Cancer Research Centre // www.bcgsc.ca



circos

CIRCOUS <http://www.circos.ca> is a genome-wide data visualization tool, developed at the University of California, Santa Cruz, to integrate genomic data, phylogenetic trees, and gene expression data. It is designed to facilitate genome-wide comparisons between genomes and in general any kind of position-invariant that relate genomic intervals. Structural variation data such as those, produced by sequence alignment and hybridization arrays, undergo comparative studies but are opaque to conventional visualization methods designed for genomic data. Compared to other tools [13-23], Circos is unique in its combination of circular data display layout, support for a wide range of data types, and its ability to support both static and dynamic, extensible, customization and automation, and maintaining a high data-to-link ratio [23] without sacrificing clarity of presentation. Circos has been used within the genomics community [4-6] and its flexibility and aesthetics has garnered interest from mainstream periodicals and newspapers [7-9] and, recently, illustrate the dynamics of a US presidential debate [10].

DOWNLOAD CHIRICOS AT <http://chiricohs.bugsy.ca/ekisles>

All present, laboratories had been greased not only to store and analyze, but to visualize the means of data produced by ultra-high-throughput technologies, such as massively parallel sequencing. Because automatically extracting meaningful patterns from these large data sets is very difficult, specialized visualization tools that highlight informative segments of the data are available in data mining and formulating hypotheses.

The design of Circos is based on the fact that a circularly composed ideogram layout can encode relationships between data elements in a way that is not possible in a linear layout. These relationships are usually represented by lines which can be either straight lines or Bézier curves whose control point location can be highly customizable. Other data types that are supported are scatter plots, pie charts, histograms, pie charts, and stacked bars. The radial position of data tracks is controlled by the user and their angular extent is a function of the extent of the data domain. Data tracks such as the tile and feed label have their individual elements automatically positioned to avoid overlap.

Circos uses **plain text files** for both **input data** and **configurations**. The latter controls the placement and format of each data track. The ability to generate both data and configuration files automatically makes Circos a very powerful tool for genome visualization.

A feature unique to Circos is not only the ability to adjust the length scale for each ideogram (e.g. display chr 17 at 2x normal size), but to **separately** vary the length scale for each

effectively zooming (or contracting) regions of interest while still displaying the entire data domain (Figures A3–A3, B1). This global and local scale adjustment is useful when visualizing genomic reports in which data density is highly variable. Furthermore, to help draw attention to important data, the ideograms can be divided into any number of distinct regions, which in turn can be drawn in any order. The resulting *ideogram tracks* can be marked up in various styles on the final image to clearly mark the distinction. Every aspect of the final image is customizable and fully *API*-able, as evidenced in a *public API* release at [NCBI](https://www.ncbi.nlm.nih.gov/ideogram/).

For example, the thickness, outline and color of the diagram track is customizable, as are the corresponding features of the cytoskeletal bands. The radial position for each diagram can be independently set. Each data track and individual primitives within a track have an associated *display value*, which controls how elements stack. Finally, every data track has a *display format*, such as *linear*, *logarithmic*, *date values*, *label* and *viscosity*. These rules can be activated by *dynamic* *formatting rules* based on *data position* and other *format values* at run-time (Figures C1-C4G). These rule sets are stored in the configuration files and separate the definition and storage of formatting rules from the raw data.

FIGURE 1

GLOBAL, HIGH-LEVEL, SELECT ADJUSTMENTS

In Figure A1, the global path (outer track) and its corresponding sectors (inner tracks) are shown. The sectors are colored by the size of the corresponding species' genome size, or the number of genes in the genome. In addition to global scales, local scales can be adjusted locally. The scale of regions of an organism can be adjusted independently of the scale of other regions. To create a continuously varying scaling factor across the entire tree, in Figure A2, a region of diagram 2 is scaled independently from the rest of the tree. This allows for the use of overlapping scales without automatically adjusting the target scale for other trees or constraints in the regions.

ADJUSTING THE SCALES OF INDIVIDUAL REGIONS

Figure A3 shows how to adjust the scale of a specific region of a diagram. A target scale is chosen, and the scale of the region is adjusted. The figure shows a zoomed-in view of a specific region, with arrows indicating the direction of adjustment. The final result is shown in Figure A4.

FIGURE 2: DATA POINTS

Figure B1 shows a circular phylogenetic tree with data points represented as small dots. The dots are colored by their position on the tree. Figure B2 shows a similar tree with data points represented as larger dots. Figure B3 shows a tree with data points represented as lines connecting the nodes. Figure B4 shows a tree with data points represented as small dots, and Figure B5 shows a tree with data points represented as large dots. Figure B6 shows a tree with data points represented as lines connecting the nodes.

FIGURE 3: DATA POINTS AS LINES

Figure C1 shows a circular phylogenetic tree with data points represented as lines. The lines connect the nodes of the tree. Figure C2 shows a similar tree with data points represented as lines. Figure C3 shows a tree with data points represented as lines. Figure C4 shows a tree with data points represented as lines. Figure C5 shows a tree with data points represented as lines.

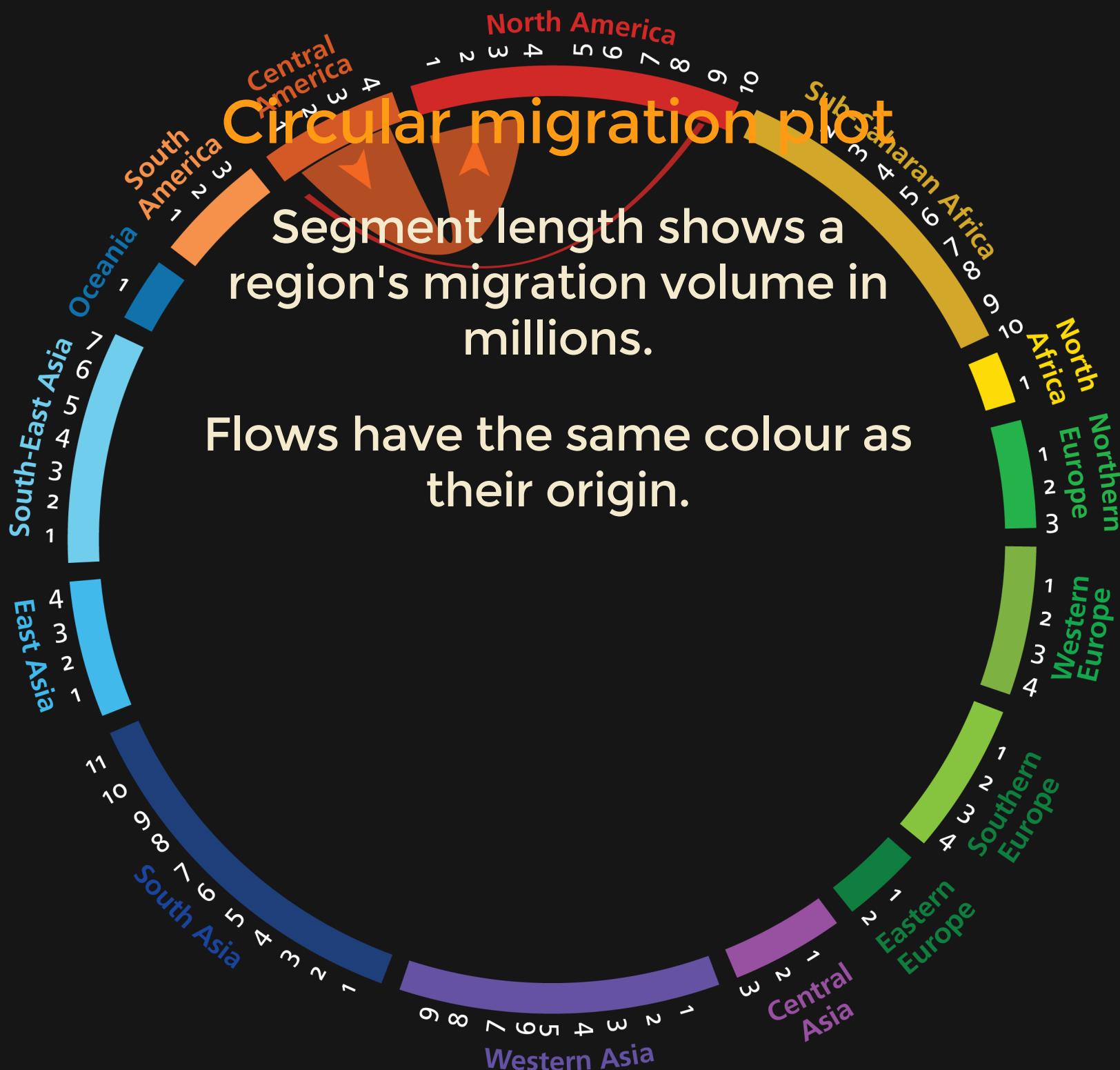
FIGURE 4: DATA POINTS AS SPHERES

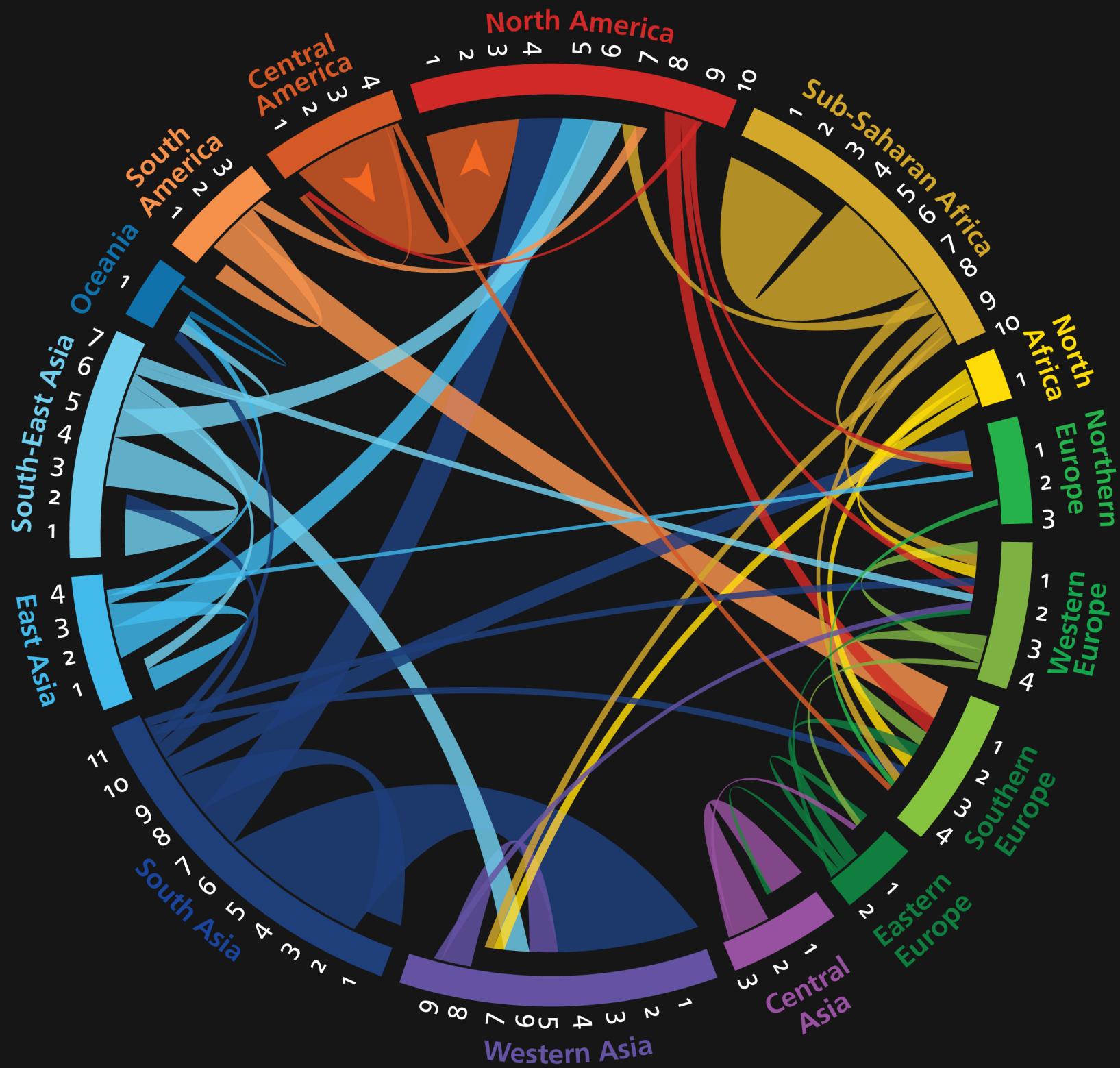
Figure D1 shows a circular phylogenetic tree with data points represented as spheres. The spheres are colored by their position on the tree. Figure D2 shows a similar tree with data points represented as spheres. Figure D3 shows a tree with data points represented as spheres. Figure D4 shows a tree with data points represented as spheres. Figure D5 shows a tree with data points represented as spheres.

FIGURE 5: DRAWING LINES

Figure E1 shows a circular phylogenetic tree with data points represented as lines. The lines connect the nodes of the tree. Figure E2 shows a similar tree with data points represented as lines. Figure E3 shows a tree with data points represented as lines. Figure E4 shows a tree with data points represented as lines.

mkweb.bcgsc.ca/circos/ // martink@bcgsc.ca





1990-1995

1995-2000

2000-2005

2005-2010



THE GLOBAL FLOW OF PEOPLE

Explore new estimates of migration flows between and within regions
for five-year periods, 1990 to 2010. Click www.global-migration.info
country-by-country.



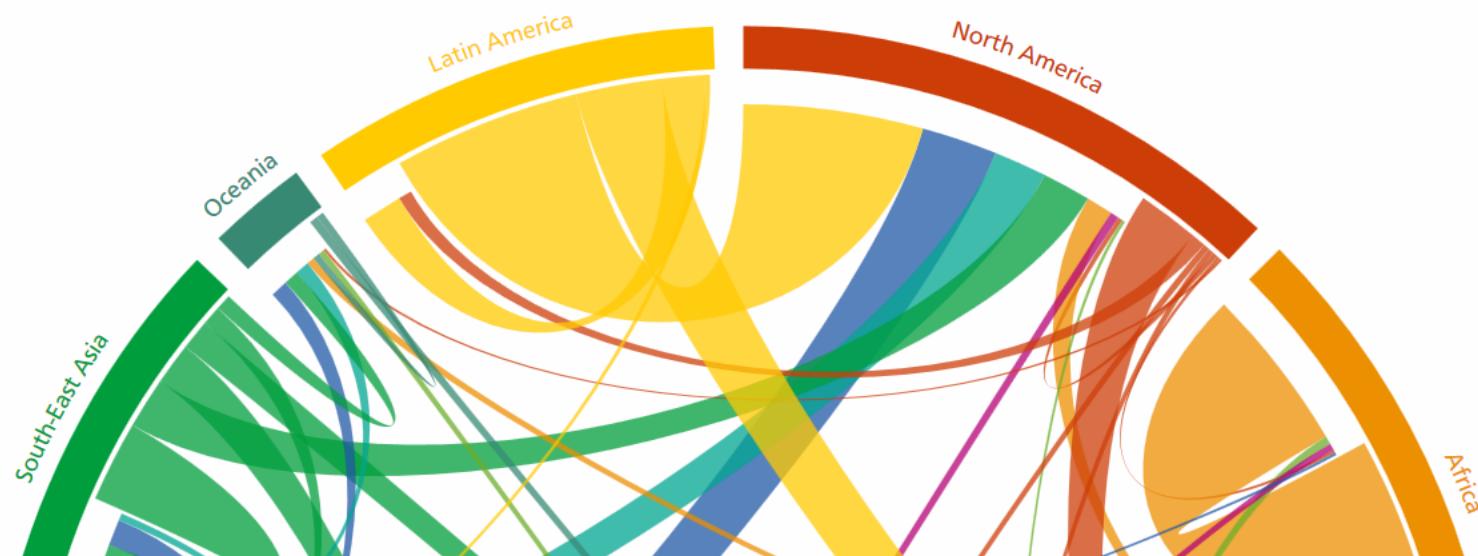
by Nikola Sander, Guy J. Abel & Ramon Bauer
at the Wittgenstein Centre for Demography and Global Human Capital

1990–1995

1995–2000

2000–2005

2005–2010



Migration in Europe

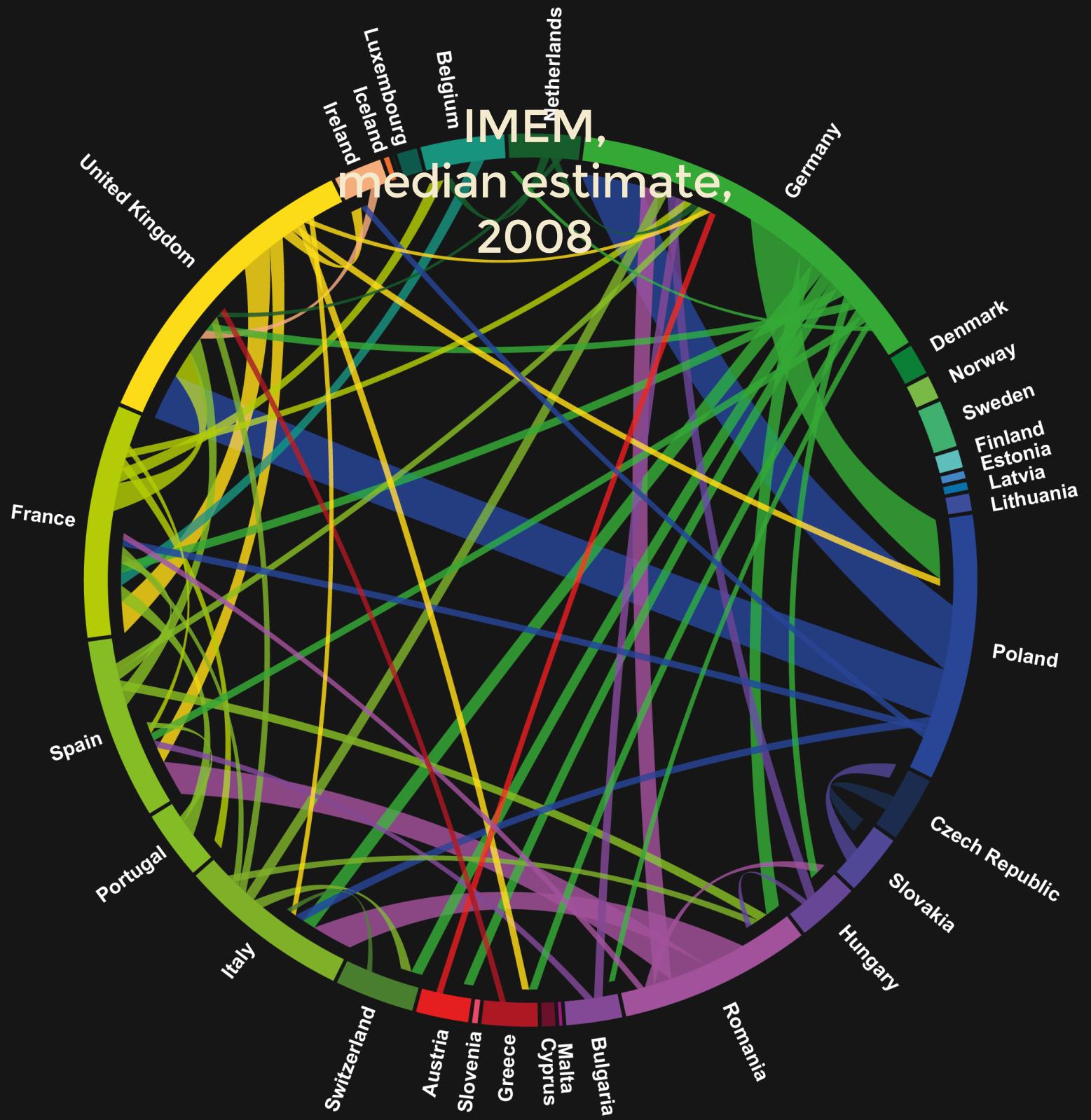
A data jungle

Population registers capture immigration and emigration through self-declaration

2 problems:

- > lack of incentives for deregistration
- > the time criterion used to identify migrants is based on intended duration of stay and varies from 3 days to 1 year

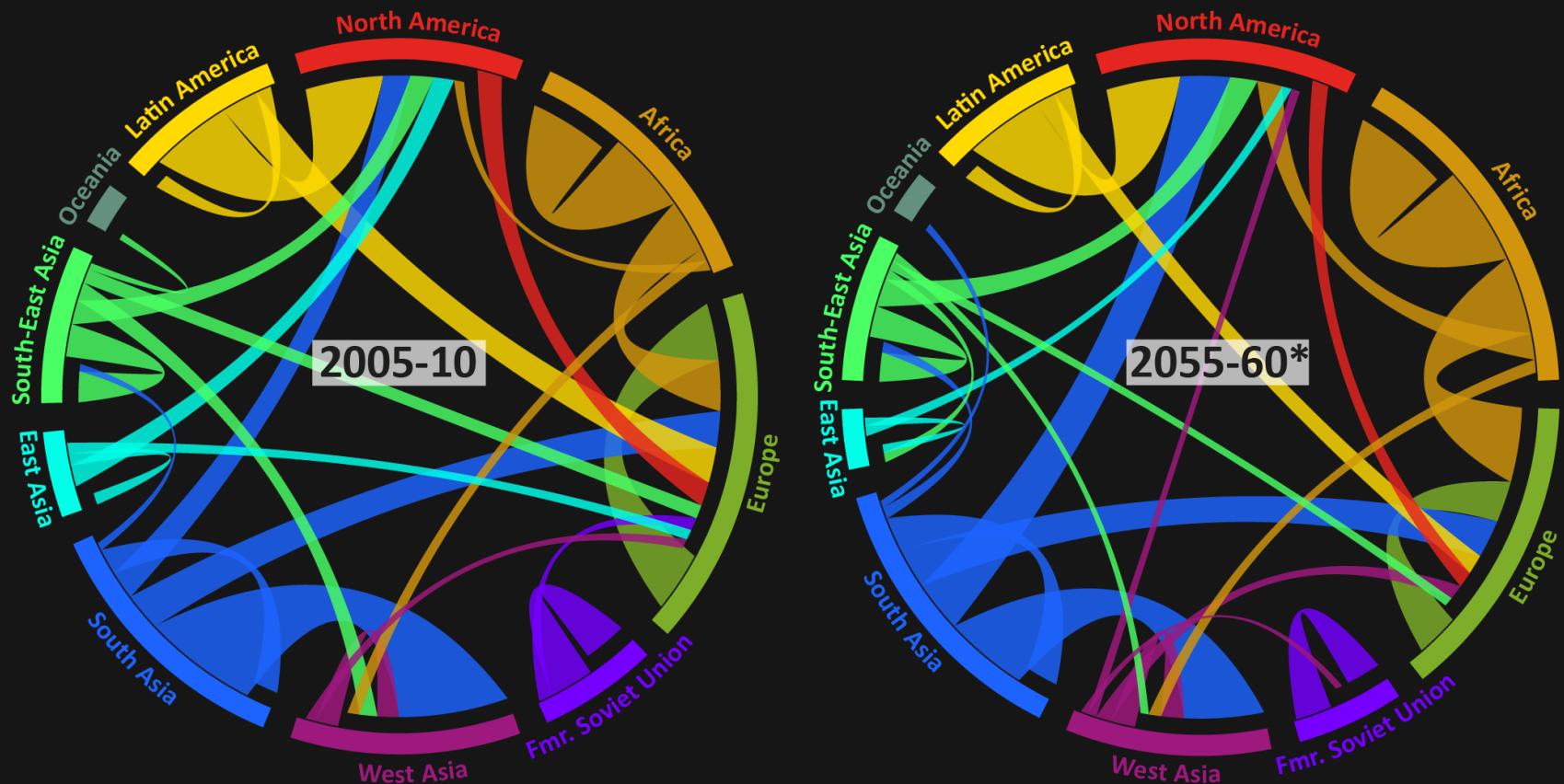
Research aims at harmonising register-based data, eg IMEM Project (Southampton)



What about the future?

Will sub-Saharan population growth
result in mass emigration to Europe?

The global migration system in 2005-10 and in 2055-60

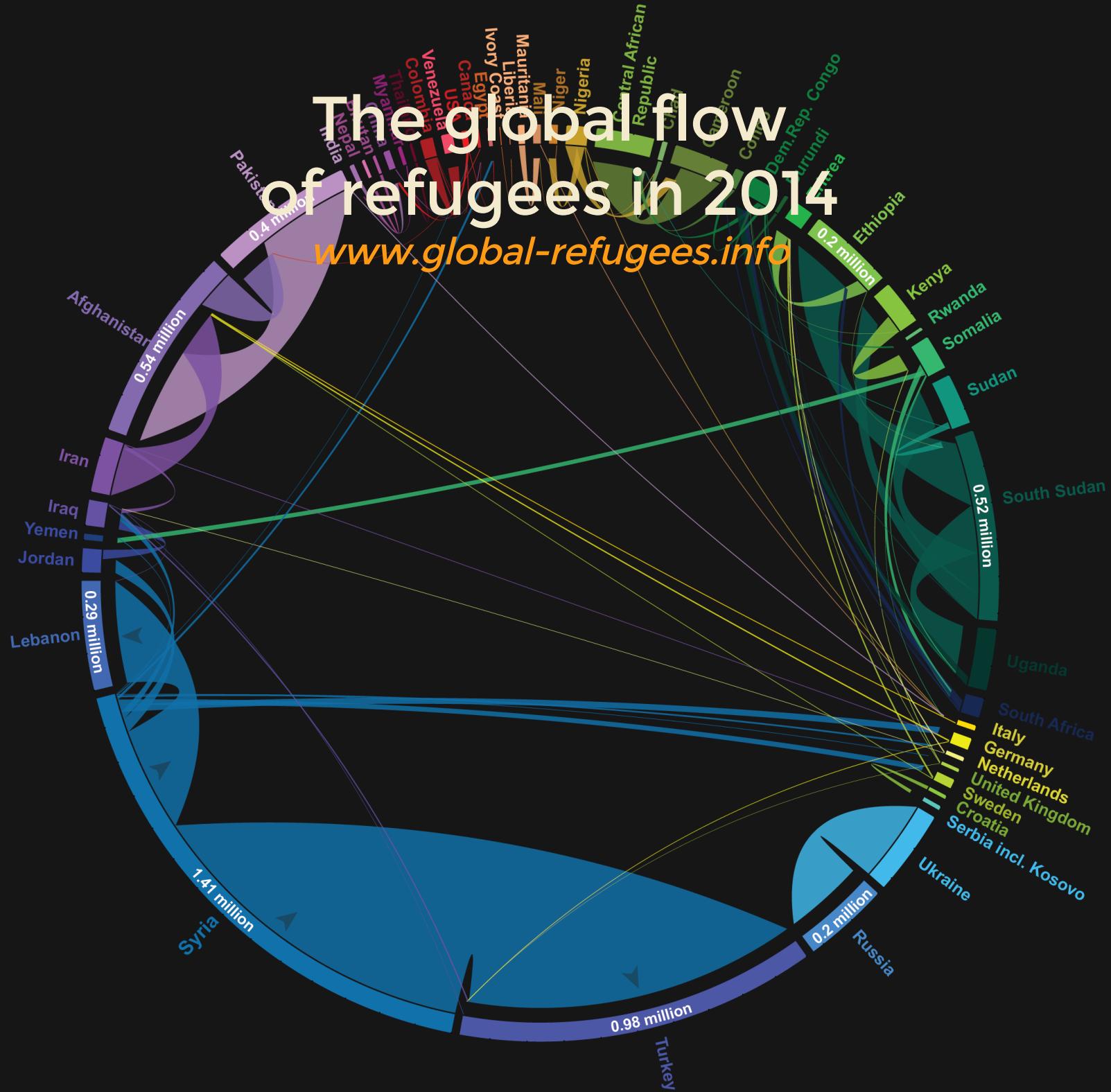


* iterative proportional fitting was used to estimate flows from projected numbers of in-migrants and out-migrants

Refugee flows triggered by violent conflict make projecting international migration difficult.

The global flow of refugees in 2014

www.global-refugees.info



Migrants or refugees?

Refugees protected by international law (1951 Convention)

Dublin III protocol

Asylum application only from inside the EU

Distribution of refugees within the EU

A common immigration policy?