

# >4.3M putative novel peptides

metasub city	sum_contig_size	protein_cnt	metasub c95 ratio	metasub c80 ratio
new_york_city	23,237,773,149	10,424,821	44.3%	8.4%
london	11,698,381,507	5,453,860	47.2%	11.0%
ilorin	12,083,747,553	5,173,133	47.6%	11.5%
(blank)	4,877,902,237	2,450,185	46.8%	12.8%
rio_de_janeiro	2,594,610,094	1,199,482	41.6%	16.8%
hong_kong	1,900,750,111	1,169,336	53.9%	9.8%
tokyo	1,862,832,603	959,027	73.5%	54.9%
bogota	1,518,739,125	778,938	56.9%	26.4%
hamilton	1,690,325,223	708,354	37.5%	7.4%
fairbanks	1,486,245,088	680,871	51.8%	19.6%
berlin	1,145,162,542	583,137	58.4%	33.5%
porto	839,222,657	443,986	66.5%	37.2%
auckland	811,476,604	352,826	47.1%	10.5%
ofa	802,189,444	345,587	41.4%	9.7%
santiago	446,643,762	252,230	63.0%	35.6%
oslo	187,649,603	96,419	49.7%	10.2%
sacramento	75,398,186	37,901	66.3%	17.1%
<b>Grand Total</b>	<b>67,259,049,488</b>	<b>31,110,093</b>	<b>47.9%</b>	<b>14.0%</b>

>1.5kb compared to JGI/NCBI, with Arbor Biotechnology, and Jonathan Gootenberg / Omar Abudayyeh

# Can we predict where you are from?

## Yes.

- R implementation of t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Perplexity 50
- $K = 2$
- Only features from Kaiju with relative abundance of at least 0.5% in at least 1 sample

