

Table S3. Difference in amino acid composition between datasets

	Bioactive long-		Non-secreted long-		Secreted long-		Bioactive long-		Non-secreted long-		Secreted long-		Bioactive short-		Non-secreted short-		Secreted short-		Avg Secreted Freq-		Avg Secreted Freq-		Uniprot Secreted-	
	Bioactive short	Non-secreted short	Non-secreted long	Secreted long	Non-secreted long	Secreted long	Non-secreted short	Secreted short	Non-secreted short	Secreted short	Non-secreted long	Secreted long	Non-secreted long	Secreted long	Non-secreted short	Secreted short	Non-secreted short	Secreted short	Bioactive long	Bioactive short	Bioactive long	Bioactive short	Uniprot non-Secreted	Secreted
A	0.2	0	0.7	1	0.1	0.1	0.5	0.6	0.1	0.1	0.5	0.6	-0.2	0	0	0	0	0	-0.2	1.2	0.6	1.2	1.2	0
C	0	0	0.2	0.7	3.5	0.1	3.3	0.6	0.1	0.1	3.3	0.6	1.2	1.2	0.6	0.6	0.6	0.6	1.2	1.2	0.6	1.2	2.4	2.4
D	0.7	0	1.2	0.3	0.5	0.2	0	0.7	0.1	0.1	0	0.7	-0.4	-0.4	0.7	0.7	0.7	0.7	-0.4	-0.4	0.7	-0.4	0	0
E	3	3	0.2	0.3	1.2	0.2	4.4	2.9	0.2	0.2	4.4	2.9	-0.1	-3.1	2.9	2.9	2.9	2.9	-0.1	-3.1	2.9	-0.1	1.2	1.2
F	3.4	3.4	0.2	0.4	0.3	0.2	3.9	4	0.2	0.2	3.9	4	0.1	3.6	4	4	4	0.1	3.6	4	0.1	3.6	0	0
G	2.5	0	0	0.6	1.2	0.4	3.6	3.5	0.4	0.4	3.6	3.5	0.1	2.6	3.5	3.5	3.5	0.1	2.6	3.5	0.1	1.3	1.3	1.3
H	0.4	0.4	0.4	0.1	0	0.2	0.1	0.4	0.2	0.2	0.1	0.4	0	0.4	0.4	0.4	0.4	0	0.4	0.4	0	0.4	0.1	0.1
I	0	0	0.5	0.1	1.4	0.1	0.8	0.3	0.1	0.1	0.8	0.3	-0.2	-0.2	0.3	0.3	0.3	-0.2	-0.2	0.3	-0.2	1.1	1.1	1.1
K	0.6	0.6	0.7	0.1	0.4	0.1	1.7	0.6	0.1	0.1	1.7	0.6	0.8	0.2	0.6	0.6	0.6	0.8	0.2	0.2	0.6	0.2	1.2	1.2
L	0.6	0.6	0.6	0.6	0.3	0.8	0.3	0.8	0.8	0.8	0.3	0.8	0.4	1.1	0.8	0.8	0.8	0.4	1.1	0.4	1.1	1.1	0.5	0.5
M	0	0	0.2	0.1	0.1	0.4	0.3	0.3	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2
N	0.5	0.5	0.5	0	0.6	1	0.6	1.4	1	1	0.6	1.4	-0.6	-1	1.4	1.4	1.4	-0.6	-0.6	1.4	-0.6	-1	0	0
P	1.3	1.3	0.2	0.2	0.2	0.3	1.1	1.3	0.3	0.3	1.1	1.3	-0.4	1	1.3	1.3	1.3	-0.4	-0.4	1.3	1	1	0.5	0.5
Q	0.7	0.7	0.7	0.6	0.6	0.4	1.9	0.5	0.4	0.4	1.9	0.5	-0.4	-1.1	0.5	0.5	0.5	-0.4	-1.1	0.5	-1.1	0.3	0.3	0.3
R	0.1	0.1	0	0.3	1.3	1.8	1.4	2.2	1.8	1.8	1.4	2.2	1.8	1.9	2.2	2.2	2.2	1.8	1.9	1.8	1.9	0.7	0.7	0.7
S	1.3	1.3	1	0.4	0.8	0.2	1.1	1.5	0.2	0.2	1.1	1.5	-0.1	-1.4	1.5	1.5	1.5	-0.1	-1.4	1.5	-1.4	0.5	0.5	0.5
T	2.1	2.1	0.2	0	0.3	0.9	2.2	3	0.9	0.9	2.2	3	-1.1	-3.3	3	3	3	-1.1	-3.3	3	-3.3	0.7	0.7	0.7
V	0.9	0.9	0.3	0.8	0.4	0.4	1	2.1	0.4	0.4	1	2.1	-0.6	-1.5	2.1	2.1	2.1	-0.6	-1.5	2.1	-1.5	0.2	0.2	0.2
W	1.1	1.1	0.1	0.1	0.2	0.2	1.4	0.8	0.2	0.2	1.4	0.8	-0.3	0.8	0.8	0.8	0.8	-0.3	0.8	0.8	0.8	0.4	0.4	0.4
Y	0	0	0.2	0.2	0.2	0.5	0.4	0.6	0.5	0.5	0.4	0.6	-0.4	-0.4	0.6	0.6	0.6	-0.4	-0.4	0.6	-0.4	0.4	0.4	0.4

Percentage differences in amino acid content between the different datasets. The first seven columns and the last column are absolute values.