

**Total  
(July 2021)**

UHGG



n = 204 398

Genbank



1 061 014

***Bacteroides  
fragilis***

757

352

RCC  
9



SRR

eLink



BioSample



BioSample

**Metadata  
annotation**

757

352

9

**Deduplication  
and quality  
control<sup>†</sup>**

382

262

4

648

**Species  
segregation**

580

68

**Removal  
overrepresented  
strains**

498\*

68\*