Last update: 08/21/2017

# GenBank Accession Number Reference Sheet

The International Nucleotide Sequence Database Collaboration (INSDC) consists of the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) and GenBank at NCBI. As part of this Collaboration, all three organizations accept new sequence submissions and share sequence data among the three databases. To facilitate the exchange of data, each member of the collaboration is assigned certain accession prefixes.

**Format for GenBank accession numbers:**

| Type | Format |
|---|---|
| Nucleotide | 1 letter + 5 numerals or 2 letters + 6 numerals |
| Protein | 3 letters + 5 numerals |
| WGS | 4 letters + 2 numerals for WGS assembly version + 6-8 numerals |
| MGA | 5 letters + 7 numerals |

**Primary GenBank accession number prefixes:**

| Accession prefixes | Data source |
|---|---|
| AE, CP, CY | Genome project data |
| U, AF, AY, DQ, EF, EU, FJ, GQ, GU, HM, HQ, JF, JN, JQ, JX, KC, KF, KJ, KM, KP, KR, KT, KU, KX, KY, MF | Direct submissions |
| AAAA-AZZZ, JAAA-JZZZ, LAAA-LZZZ, MAAA-MZZZ, NAAA-NZZZ, PAAA-PZZZ, QAAA-QZZZ, RAAA-RZZZ | Whole genome shotgun sequences |
| AAA-AZZ | Protein ID |
| EAA-EZZ, KAA-KZZ, OAA-OZZ | WGS protein ID |

**Version number suffix:**

GenBank sequence identifiers consist of an accession number of the record followed by a dot and a version number (i.e. accession.version). The version number will increment by one when there is an update to the sequence record.

**Format for RefSeq accession numbers:**

RefSeq accession numbers do not follow the naming conventions set by INSDC, they have a two-letter prefix followed by an underscore. RefSeq records are classified as “Known RefSeq” (manually reviewed by NCBI staff or collaborators) or “Model RefSeq” (records produced by an automated annotation pipeline).

| Accession prefixes | Type | Description |
|---|---|---|
| NC_, NG_ | Known RefSeq | Genomic regions or assembly |
| NM_ | Known RefSeq | mRNA |
| NR_ | Known RefSeq | Non-coding RNAs |
| NP_ | Known RefSeq | Protein |
| NT_, NW_ | Known RefSeq | Genomic contig or scaffold |
| XM_ | Model RefSeq | mRNA |
| XR_ | Model RefSeq | Non-coding RNAs |
| XP_ | Model RefSeq | Protein |

The complete list of accession numbers is available at https://www.ncbi.nlm.nih.gov/Sequin/acc.html.