Genbank file validation checks/fixes

- Verify standard fields are present
  - LOCUS, ACCESSION, etc.
- Shorten locus name to <16 characters using accession ID if possible
- Ensure Accession ID did not start with a number
  - Append "ID" to beginning if it did
- Ensure sequence is a DNA alphabet
- Replace non-alphanumeric chars with "_"
  - DEFINITION, SOURCE, ORGANISM, LOCUS fields
  - FEATURES.source organism and strain tags

Genbank validation (cont.)

- Replace all non-ACGT characters in sequence with “N”
- Check for presence of gene annotation features
- Fix invalid db_xref qualifiers
- Remove DNA coords from circular DNA that overlapped from end to beginning of sequence
  - This fixed bugs in future pipeline scripts