
QIAGEN

Table of contents

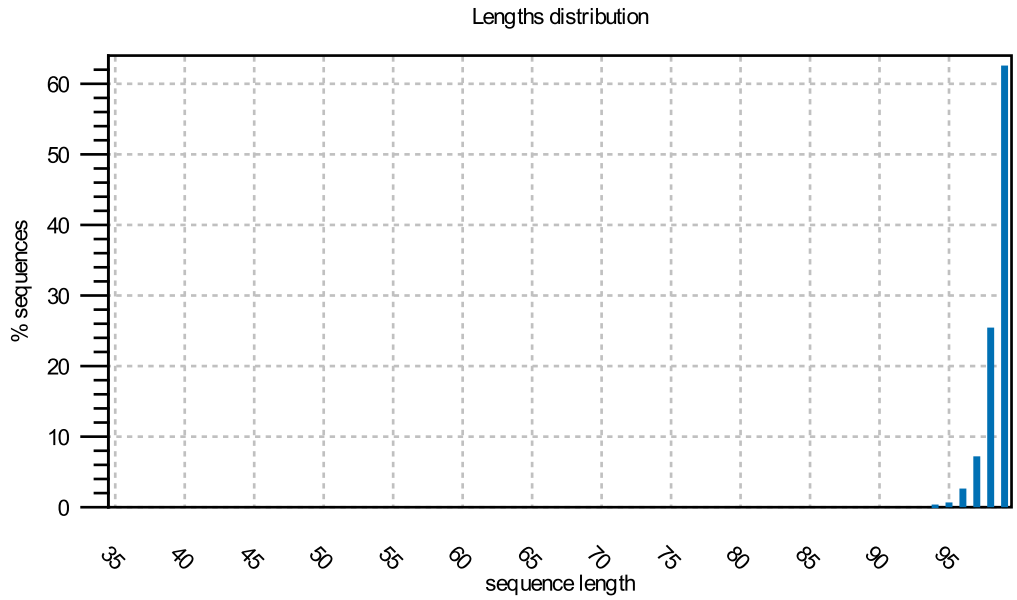
1. Summary 3
2. Per-sequence analysis 3
    2.1 Lengths distribution 3
    2.2 GC-content 4
    2.3 Ambiguous base-content 4
    2.4 Quality distribution 5
3. Per-base analysis 5
    3.1 Coverage 6
    3.2 Nucleotide contributions 6
    3.3 GC-content 7
    3.4 Ambiguous base-content 7
    3.5 Quality distribution 8
4. Over-representation analyses 8
    4.1 Enriched 5-mers 9
    4.2 Sequence duplication levels 9
    4.3 Duplicated sequences 10

1. Summary

Creation date:	Fri Dec 23 11:58:52 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
NL1_S1:	71,177,957 sequences
Total sequences in data set	71,177,957 sequences
Total nucleotides in data set	6,997,952,203 nucleotides

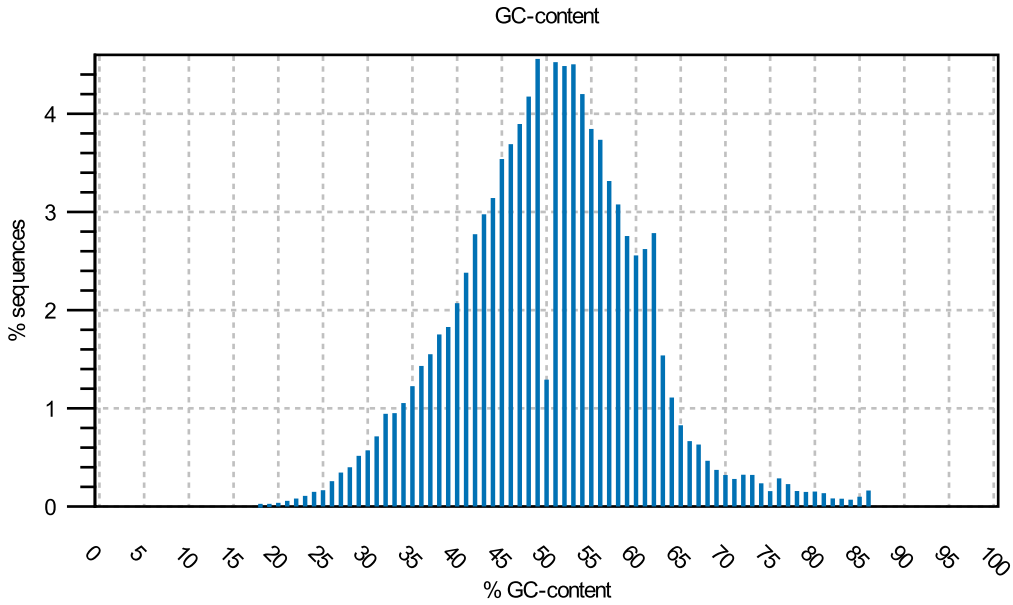
2. Per-sequence analysis

2.1 Lengths distribution



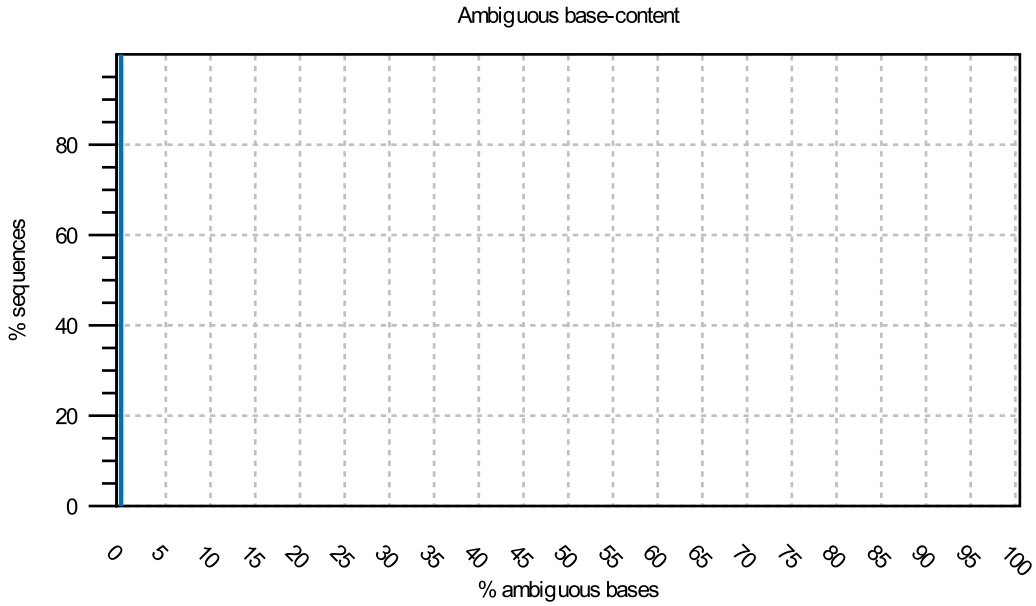
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



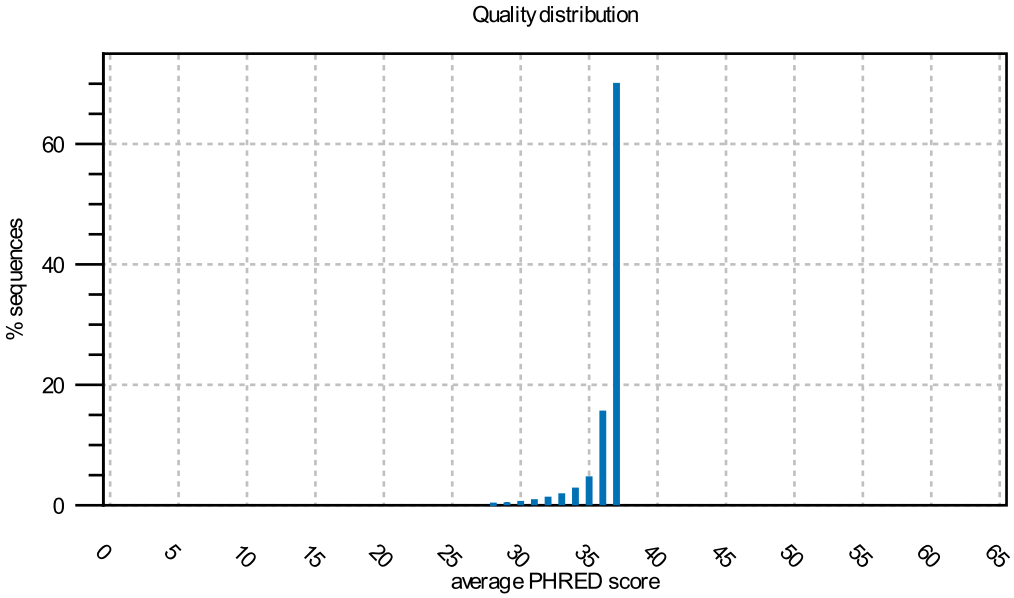
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

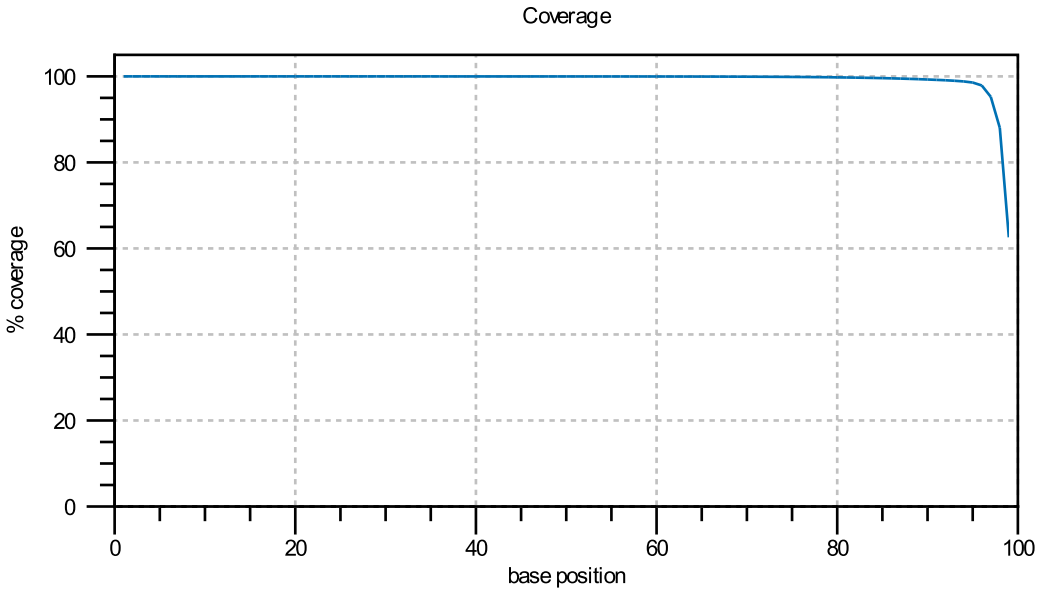
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

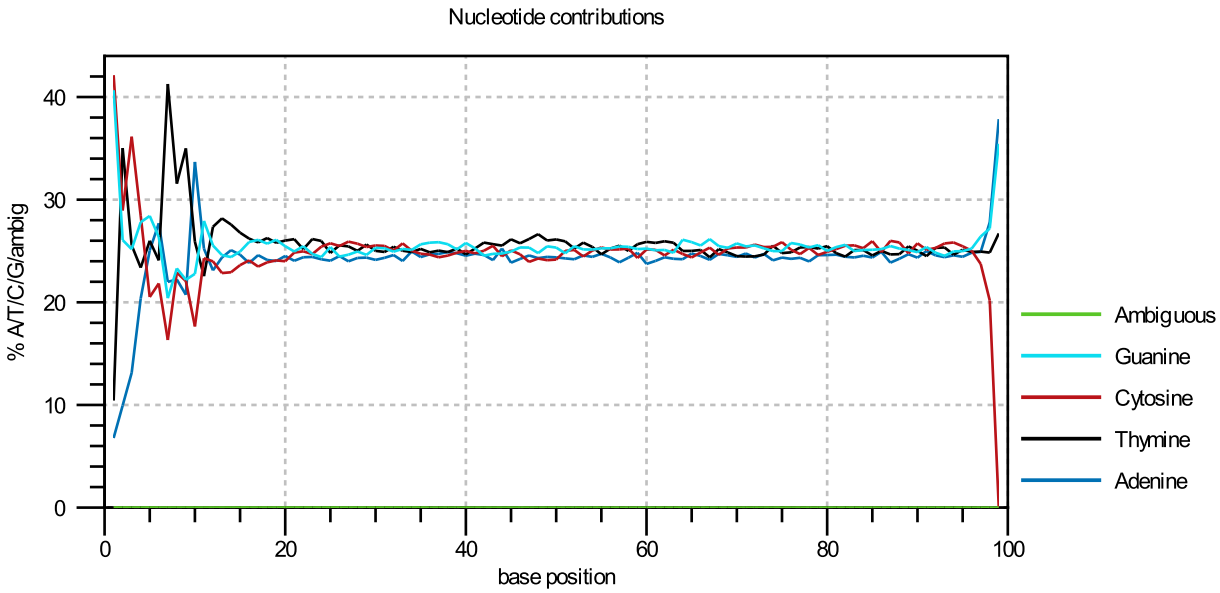
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

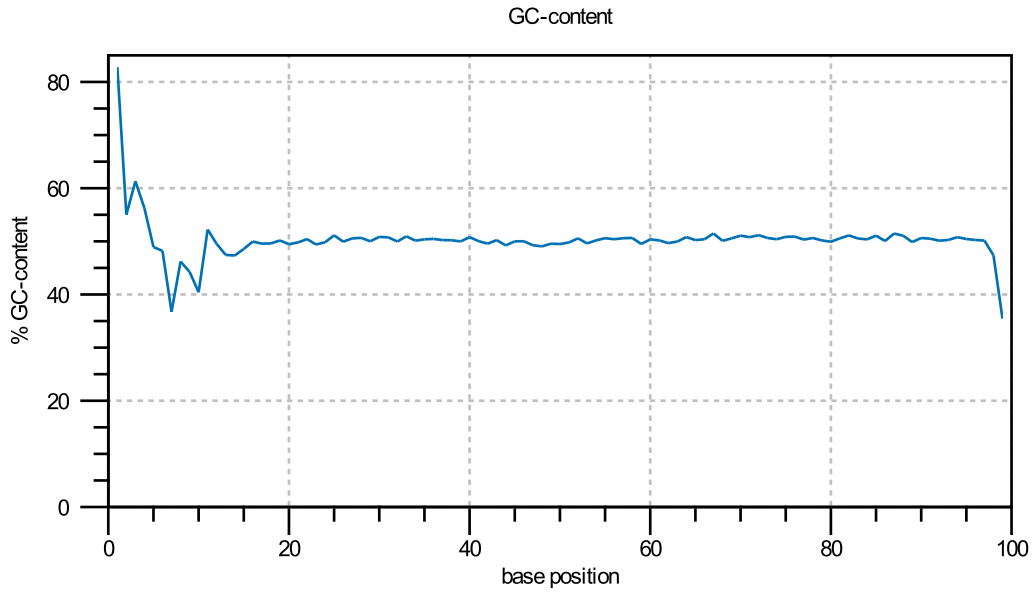
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

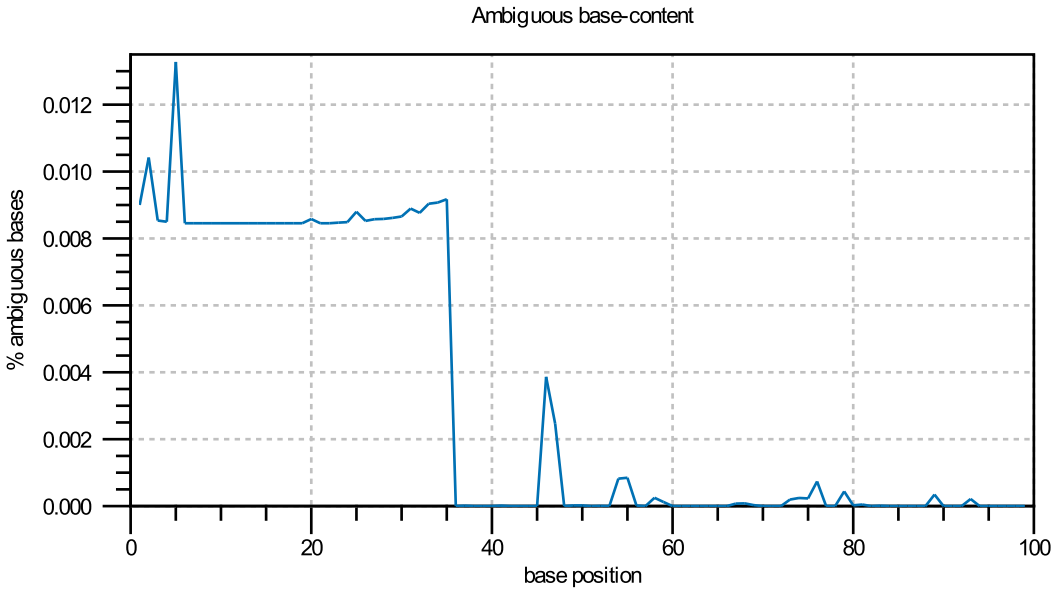
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



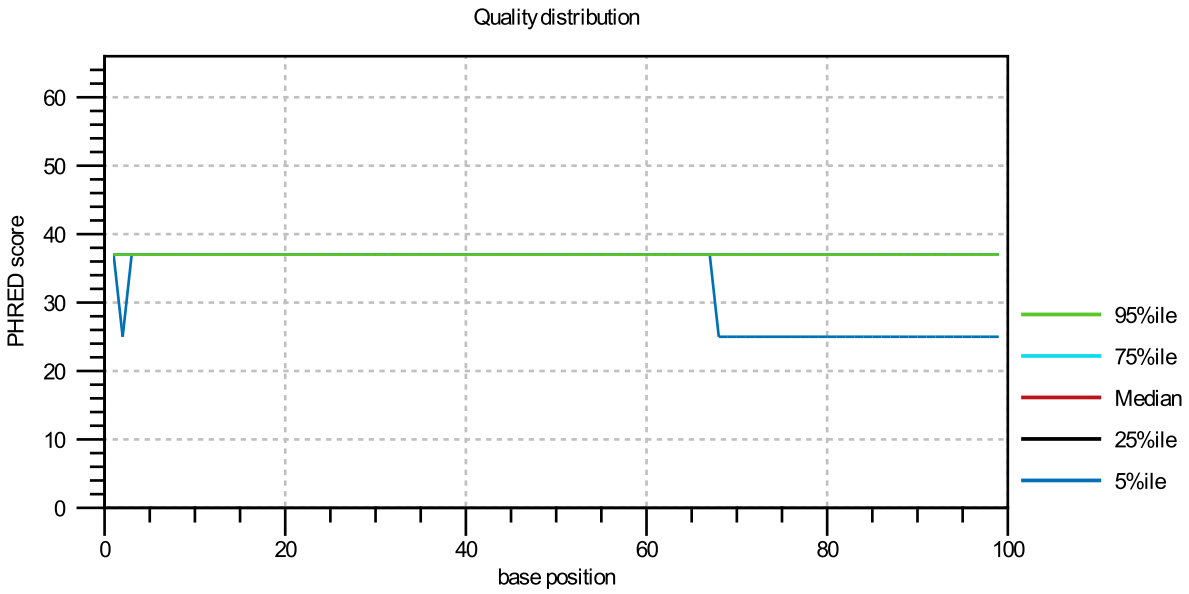
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

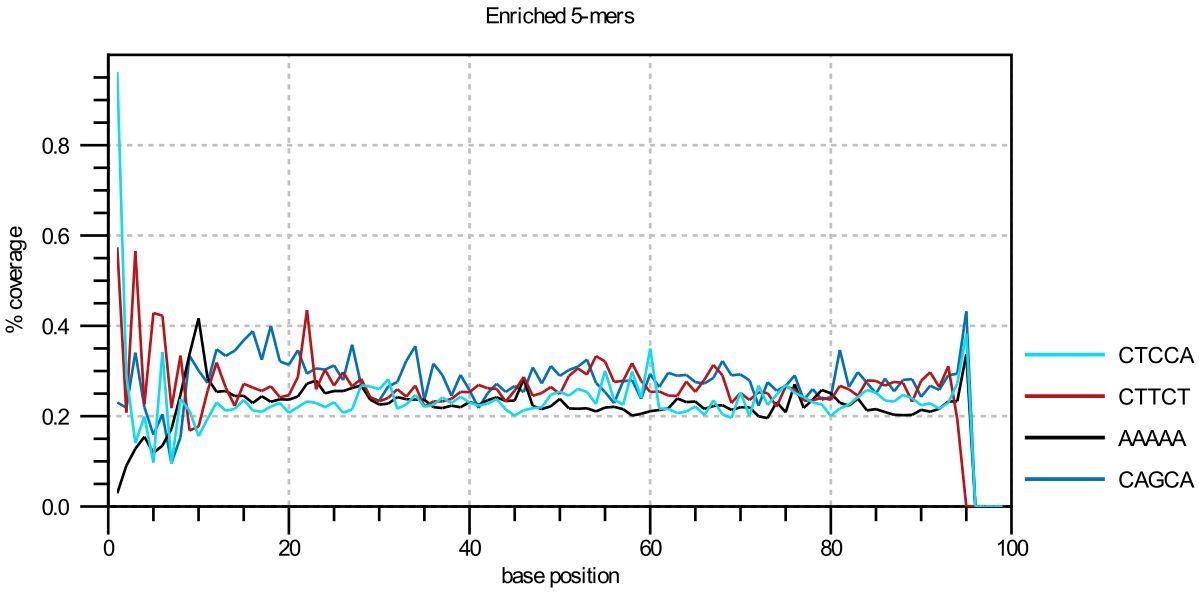
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

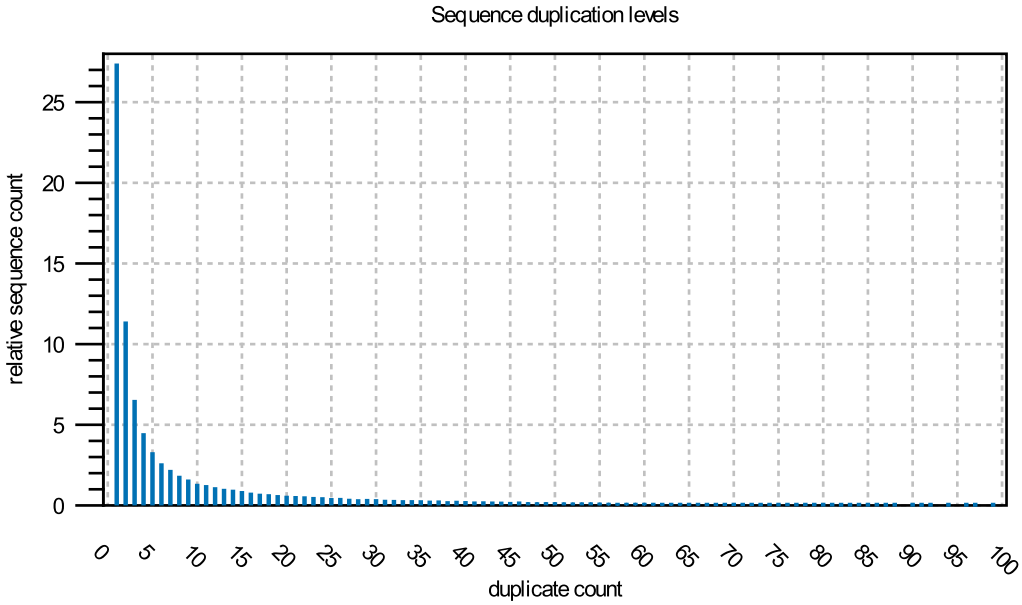
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

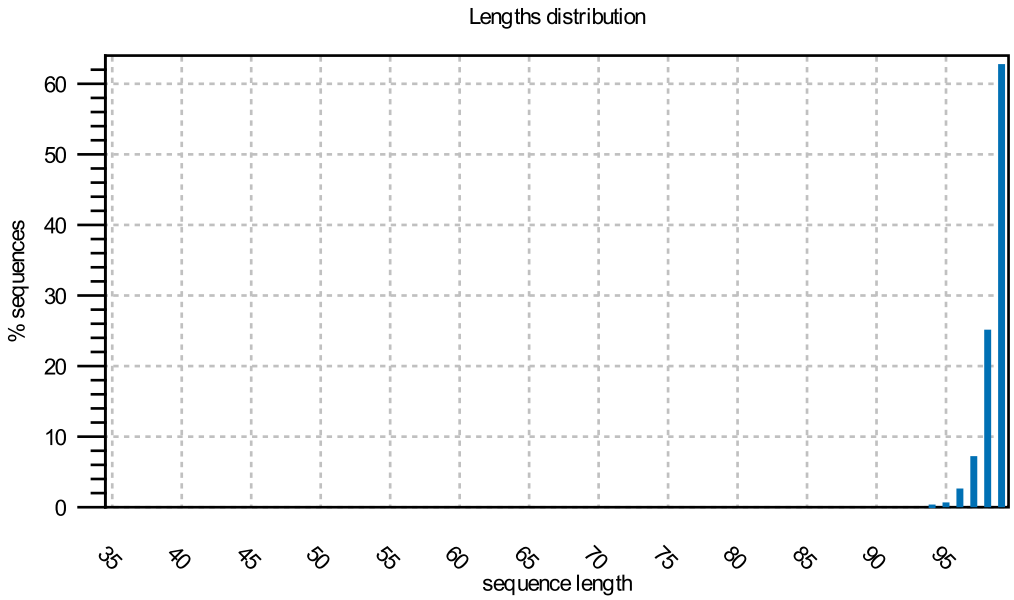
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 13:47:25 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
NL2_S11:	78,295,601 sequences
Total sequences in data set	78,295,601 sequences
Total nucleotides in data set	7,697,408,258 nucleotides

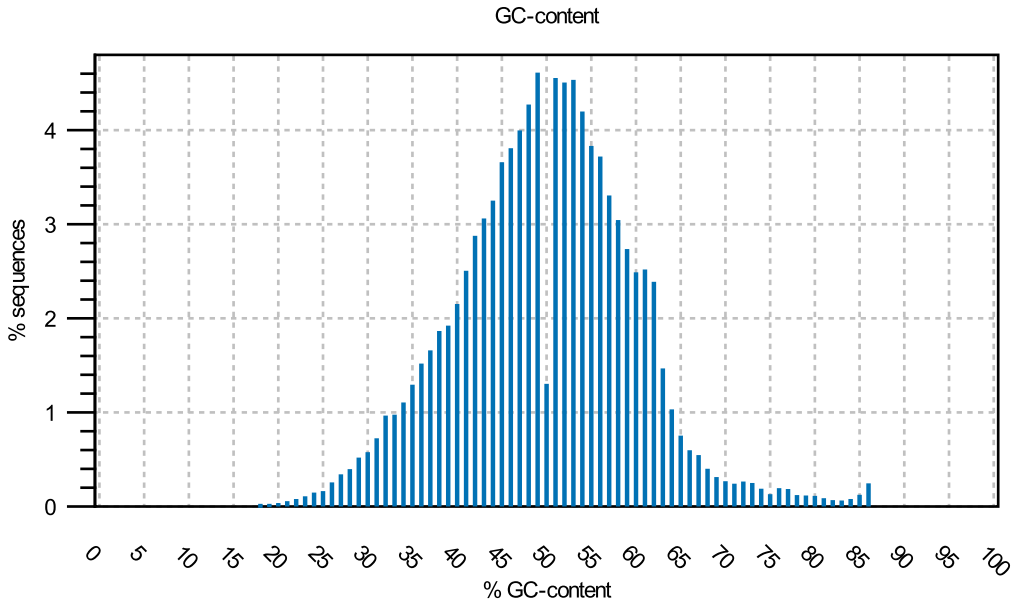
2. Per-sequence analysis

2.1 Lengths distribution



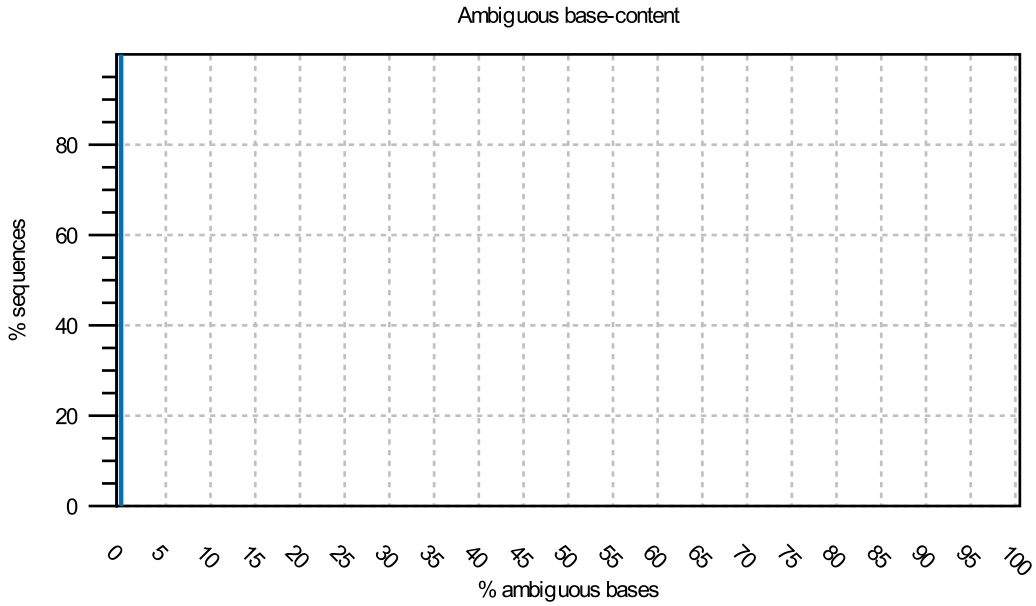
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



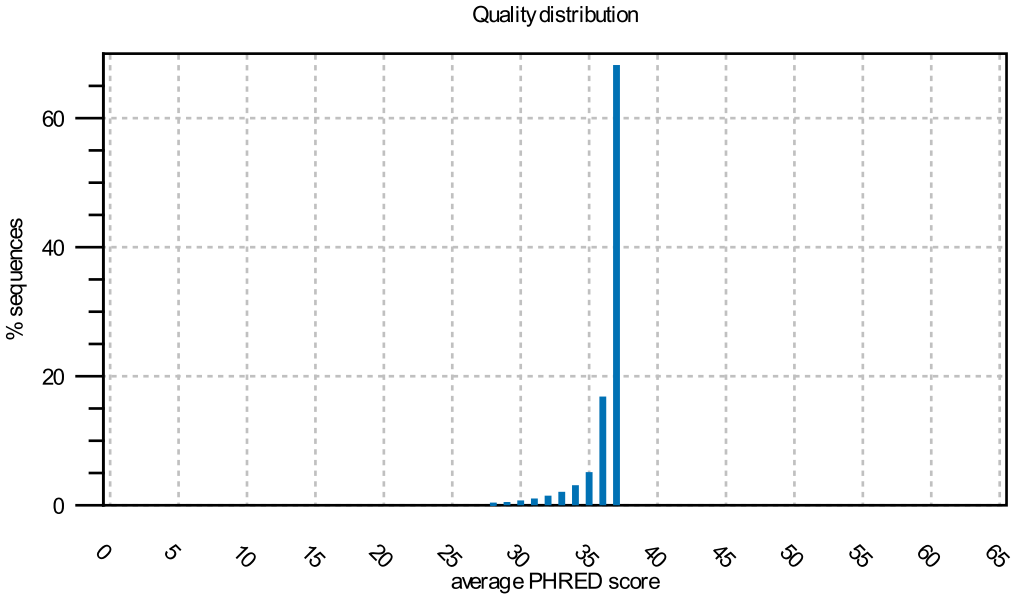
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

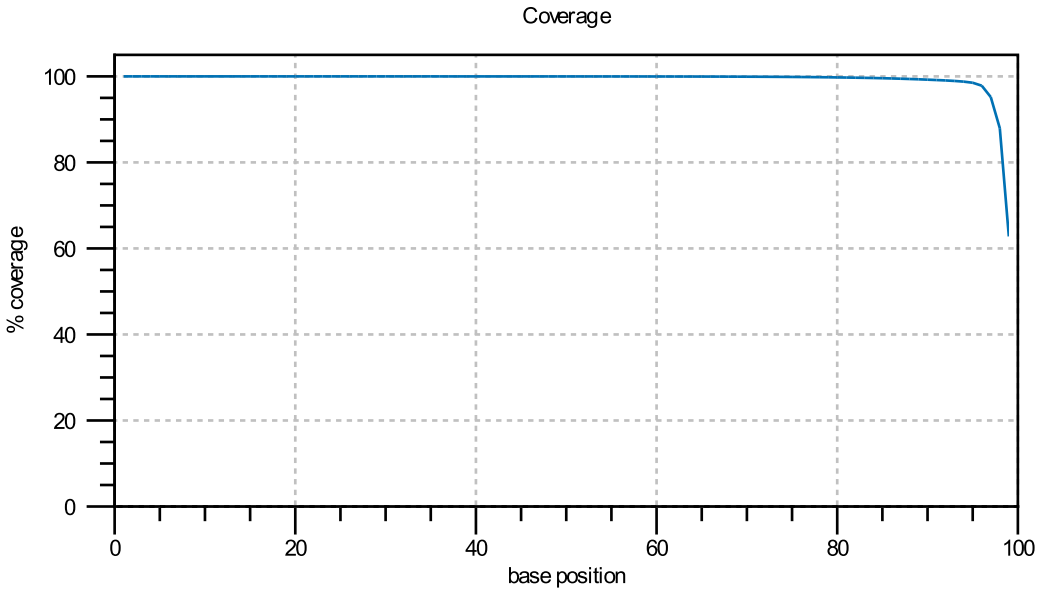
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

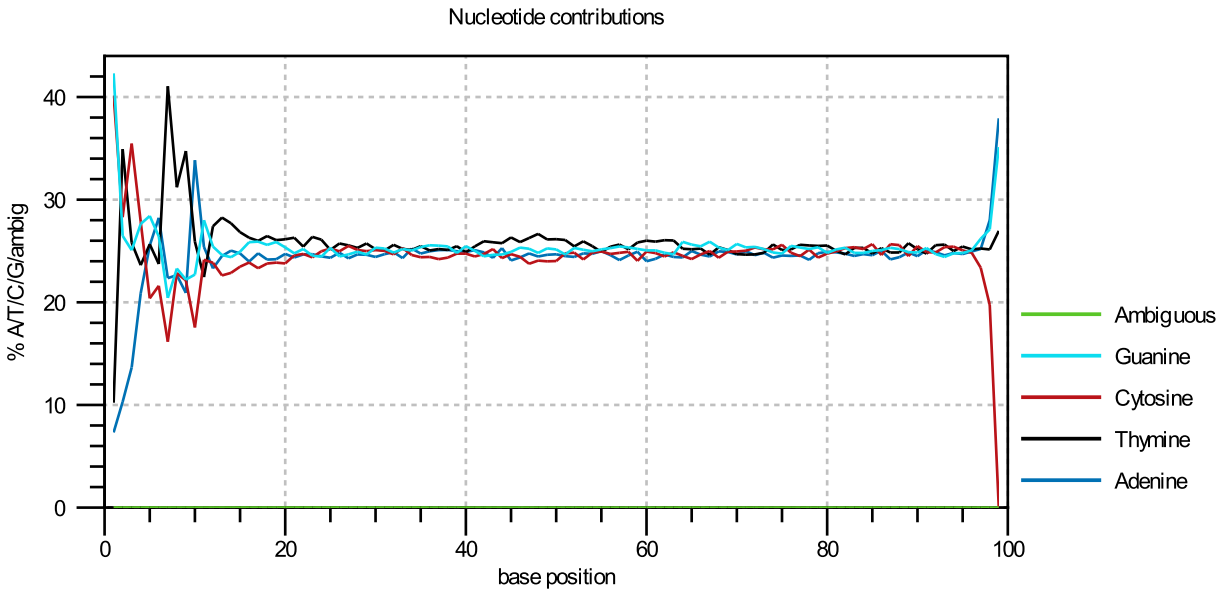
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

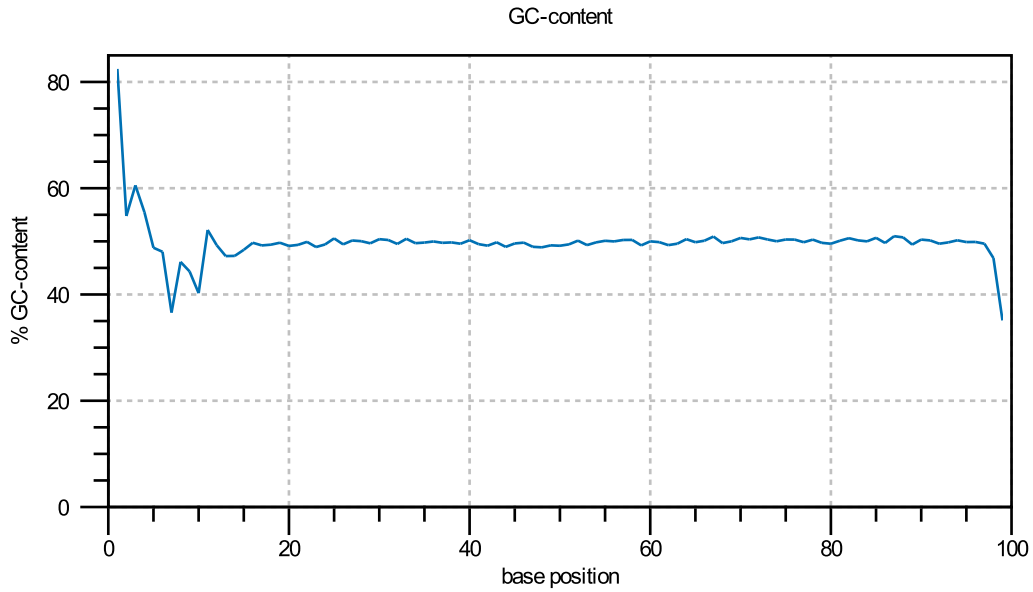
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

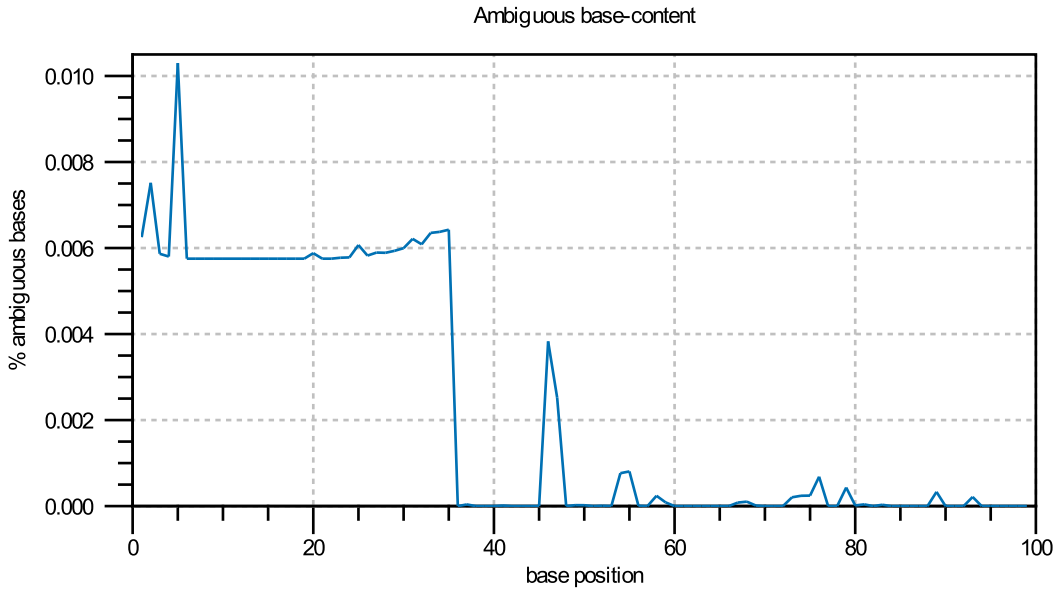
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



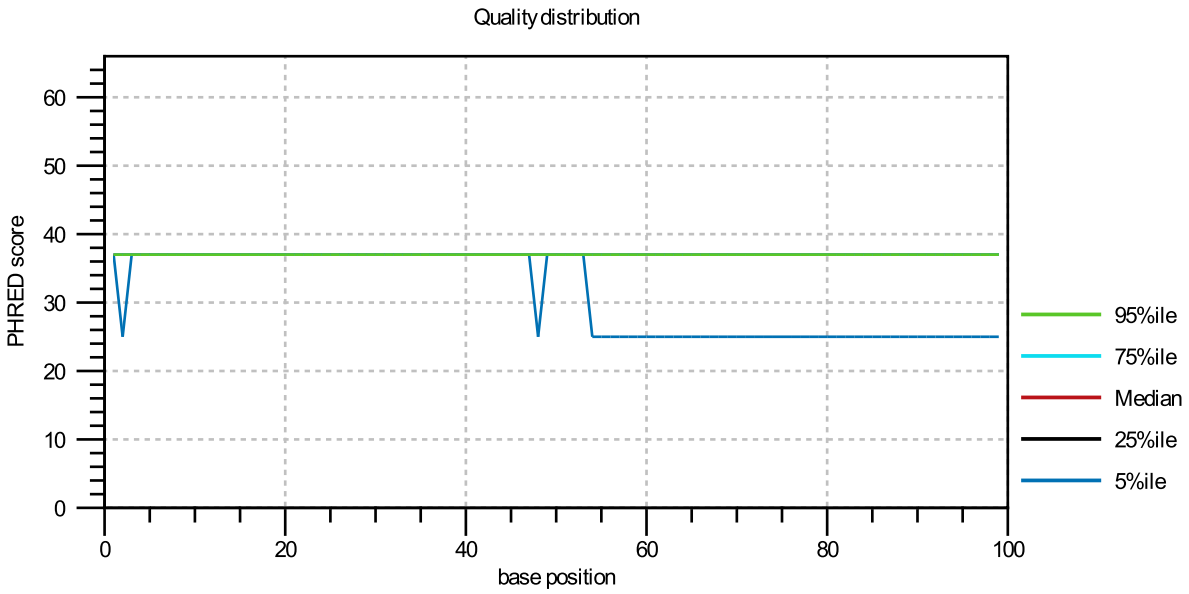
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

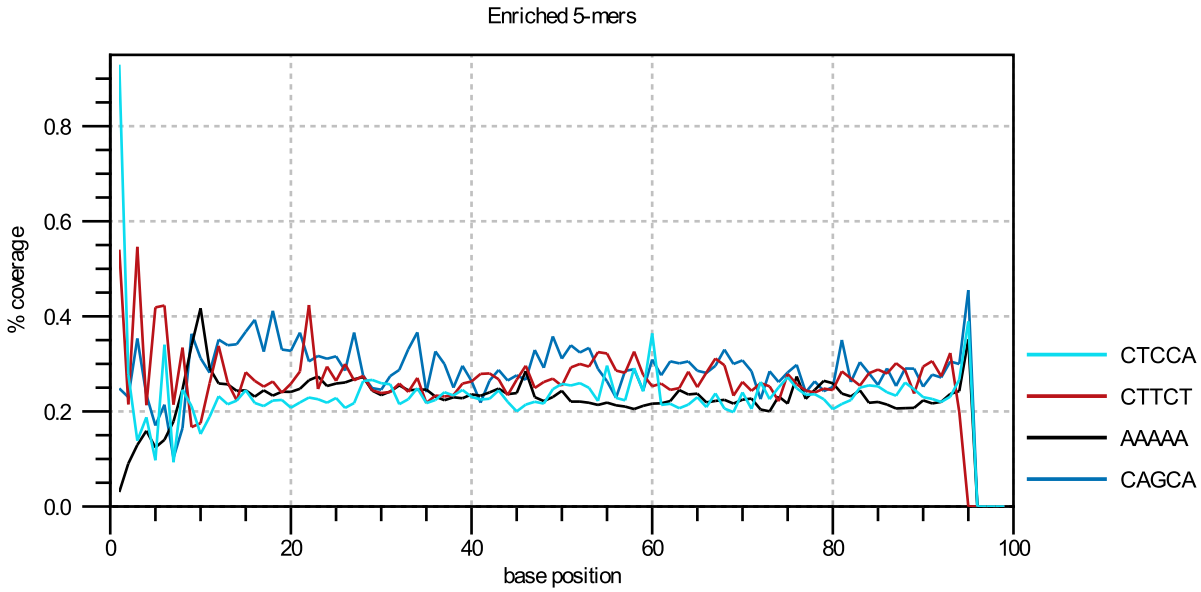
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

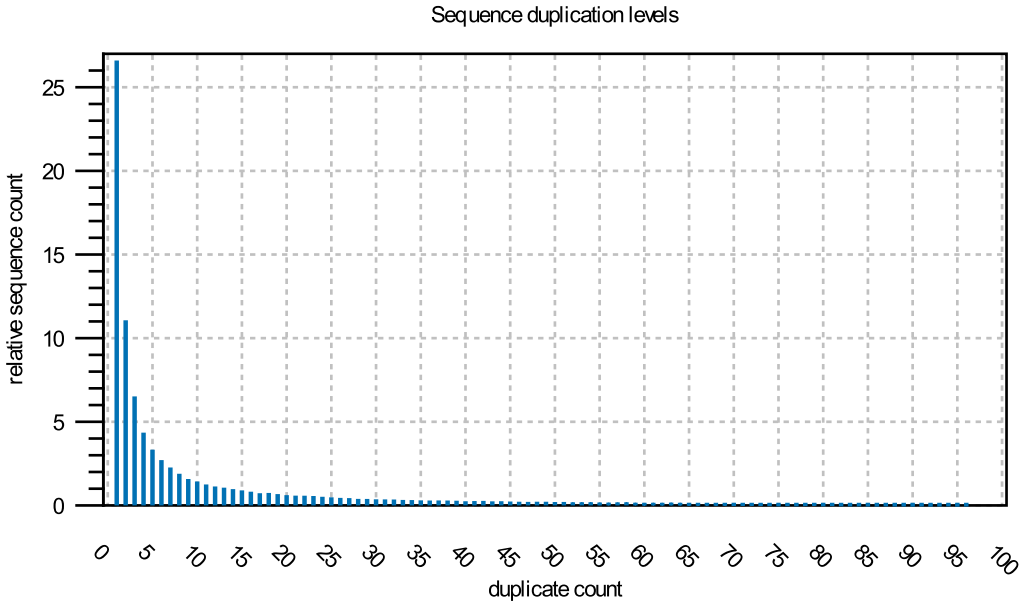
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

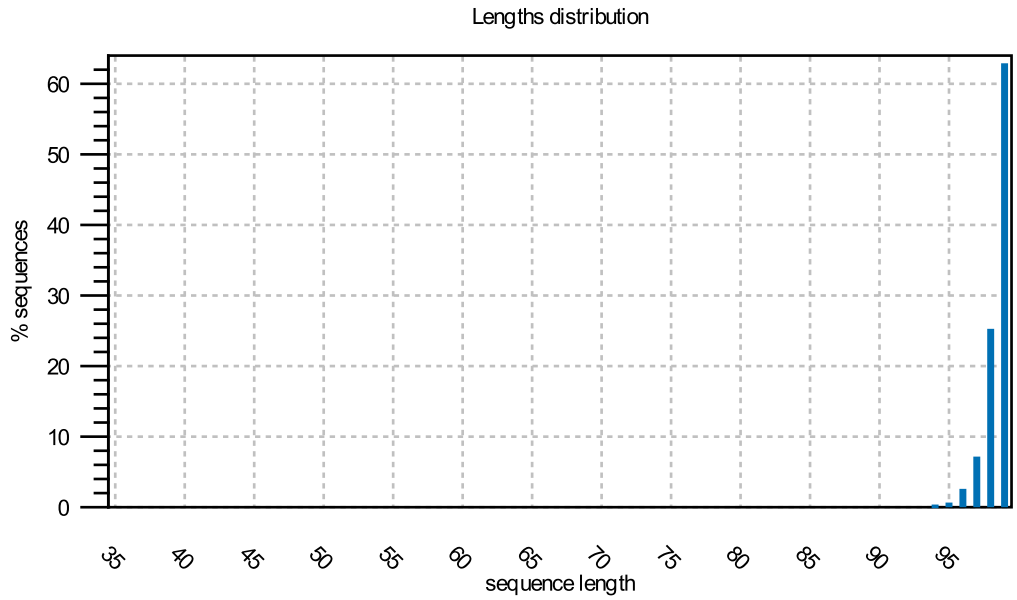
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 14:08:20 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
NL3_S4:	75,867,324 sequences
Total sequences in data set	75,867,324 sequences
Total nucleotides in data set	7,459,702,237 nucleotides

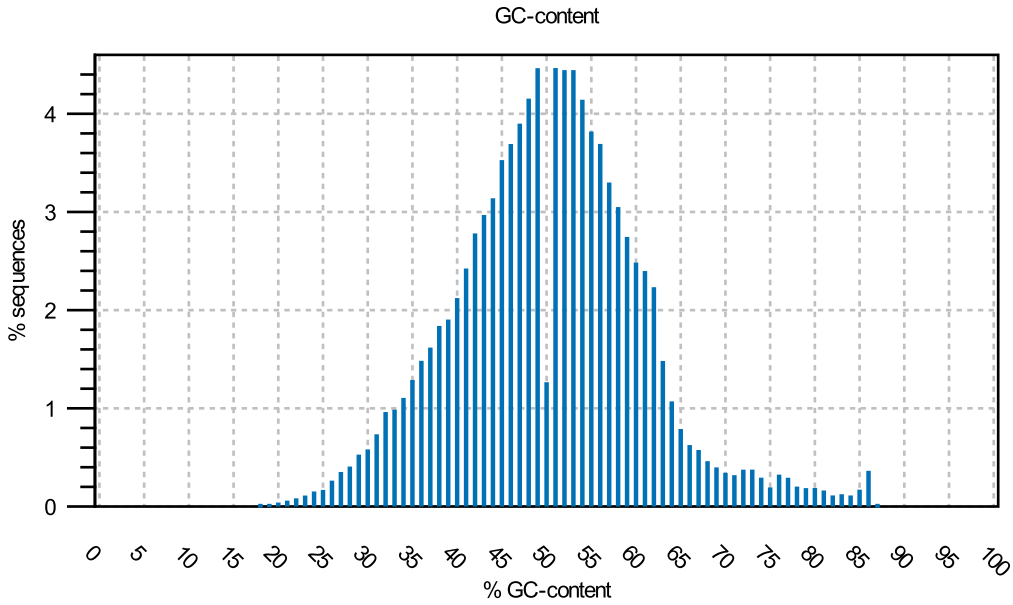
2. Per-sequence analysis

2.1 Lengths distribution



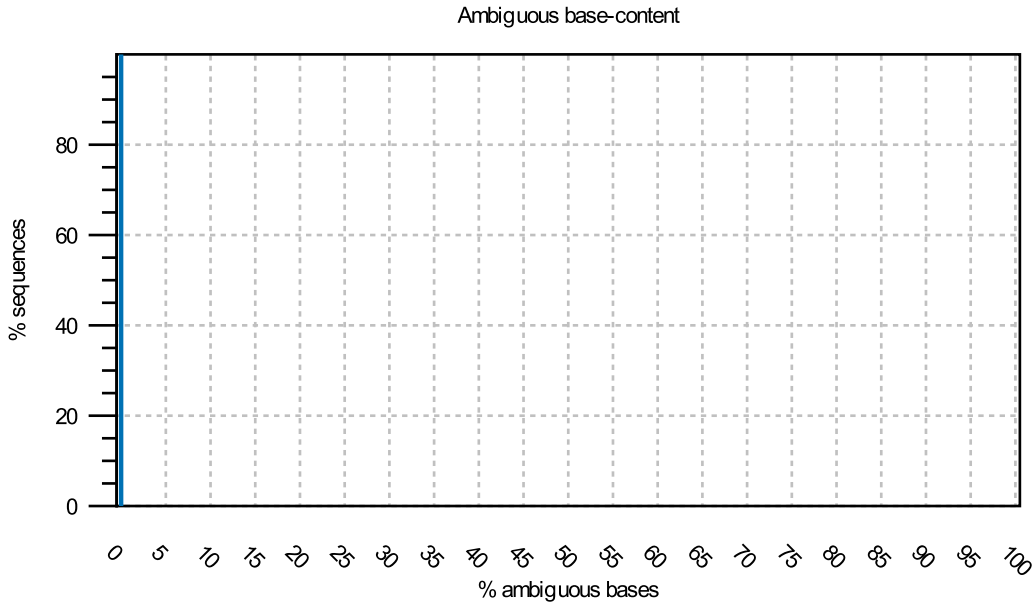
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



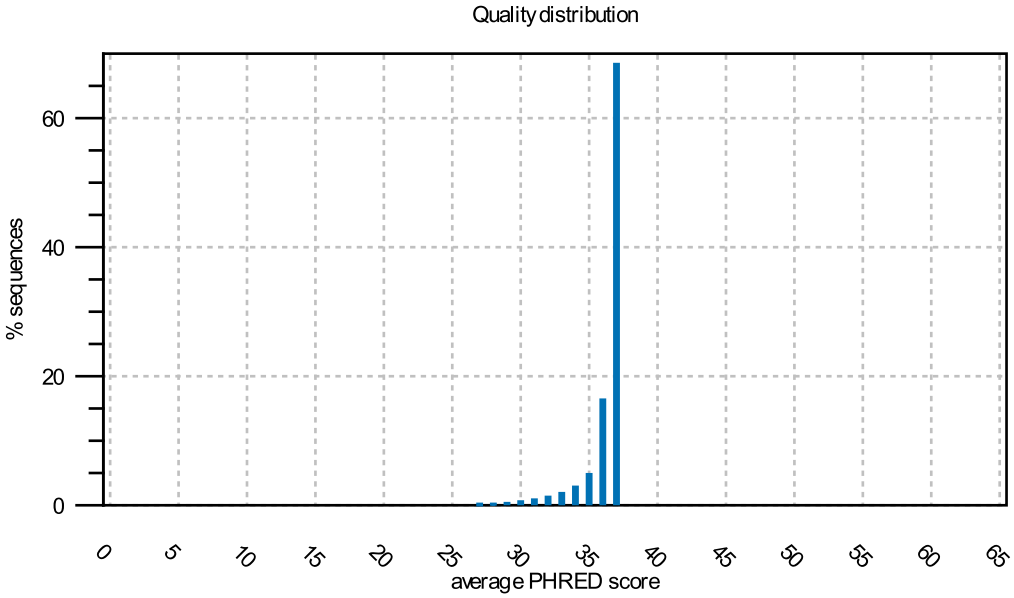
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

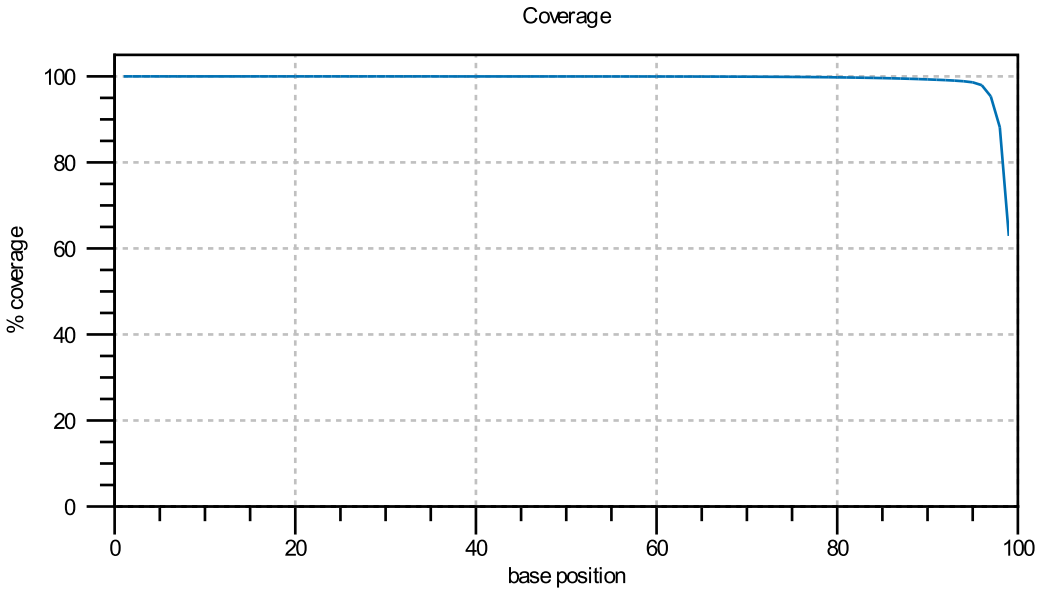
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

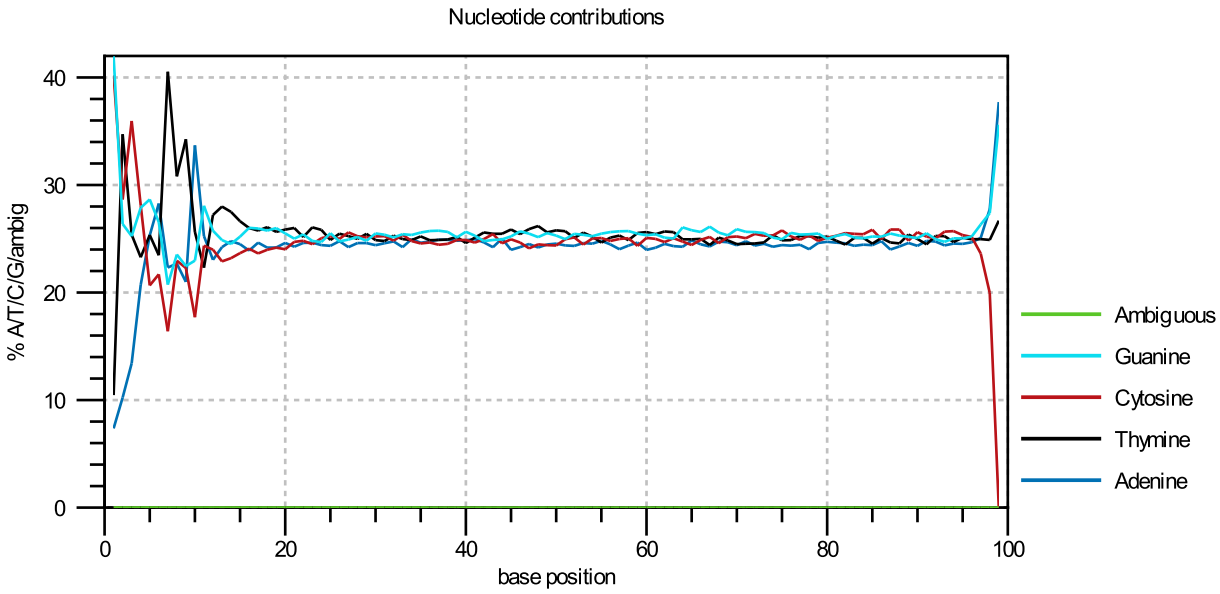
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

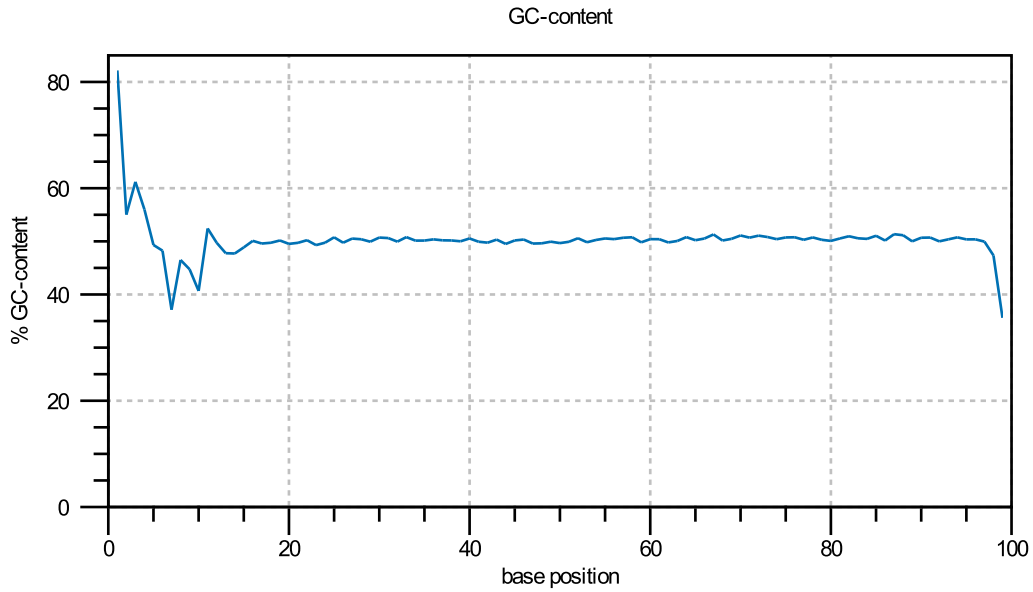
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

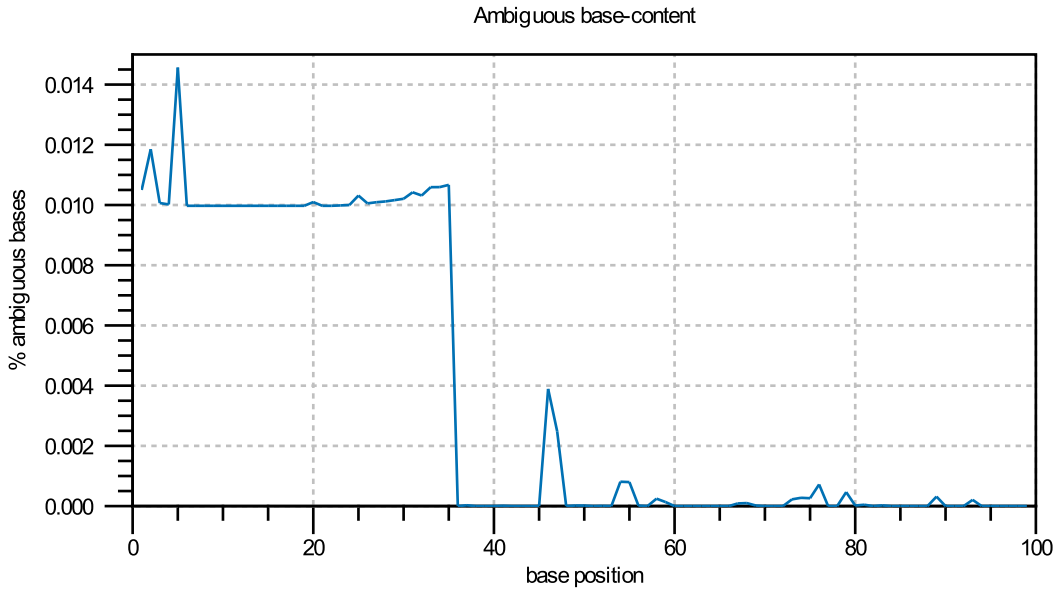
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



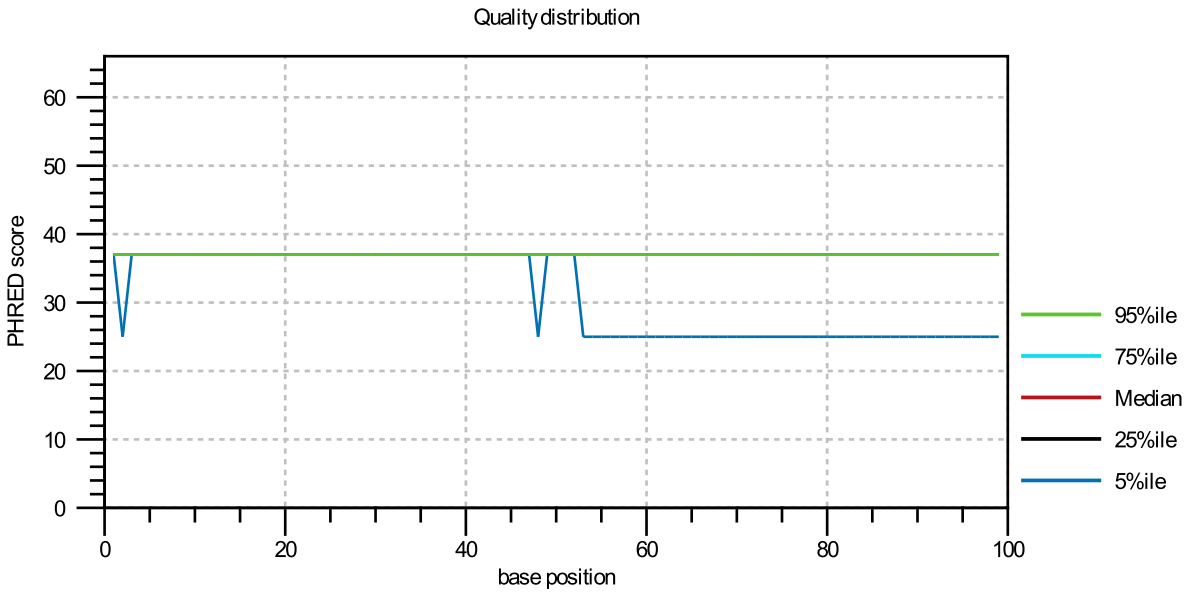
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

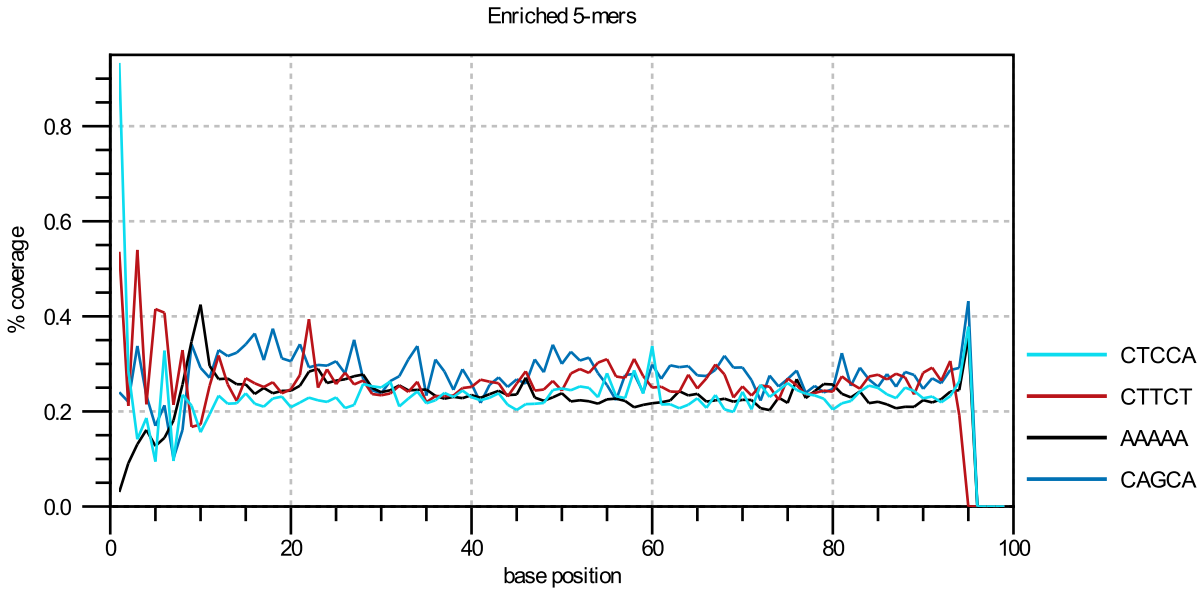
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

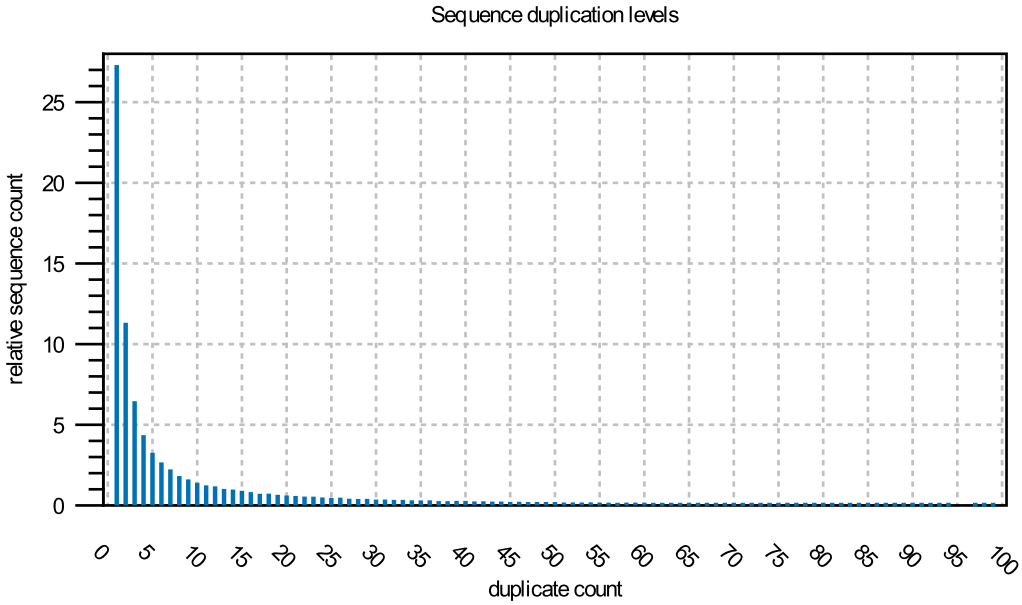
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

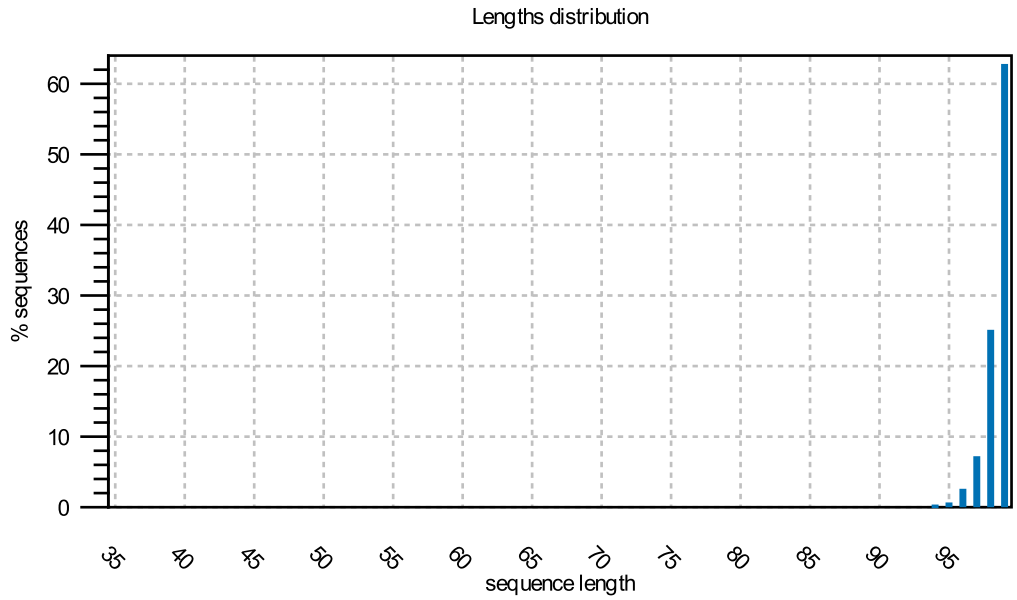
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 14:30:19 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
DL1_S7:	71,673,917 sequences
Total sequences in data set	71,673,917 sequences
Total nucleotides in data set	7,045,249,385 nucleotides

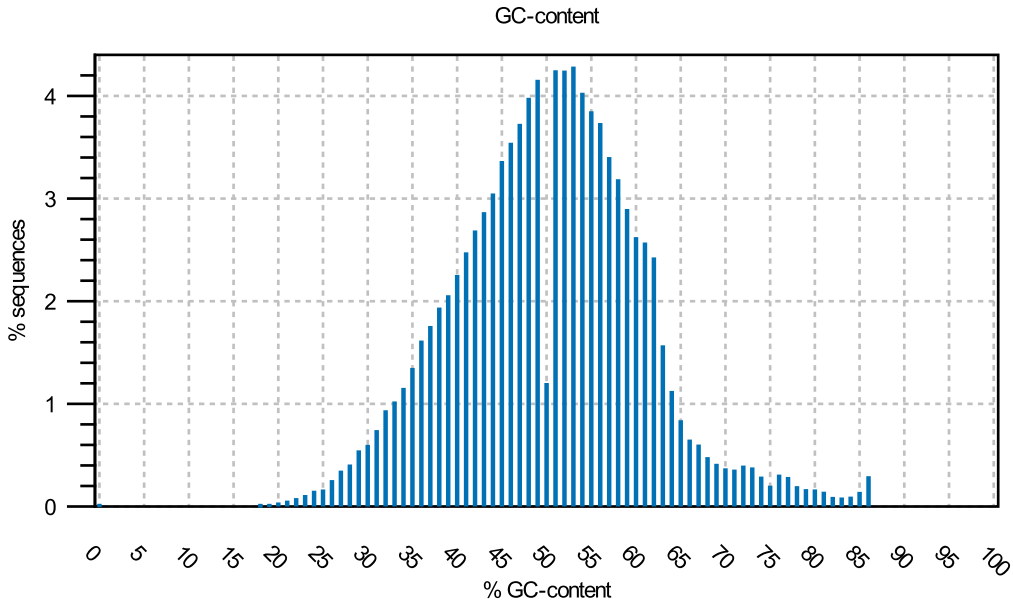
2. Per-sequence analysis

2.1 Lengths distribution



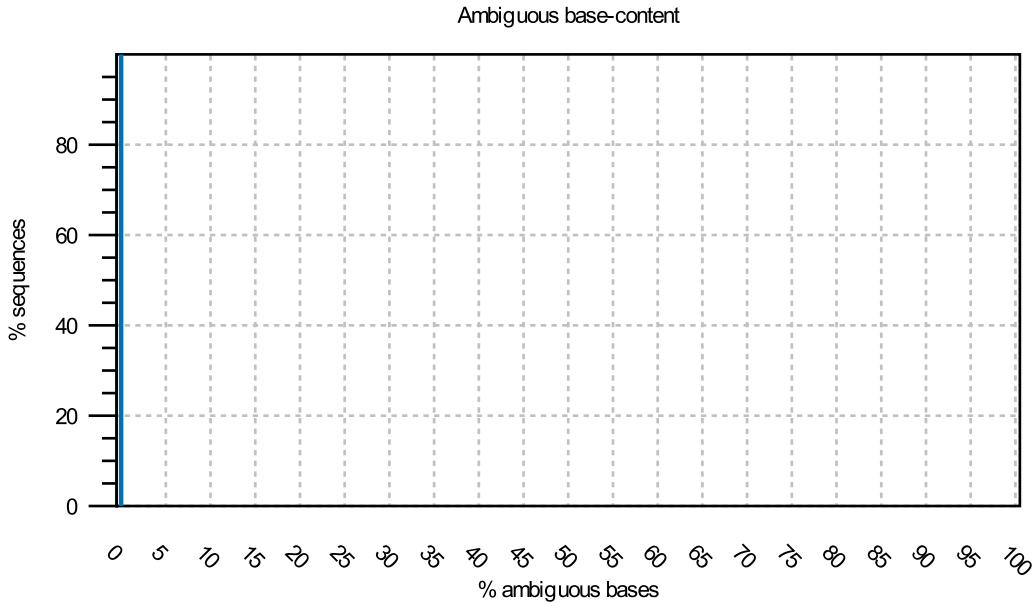
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



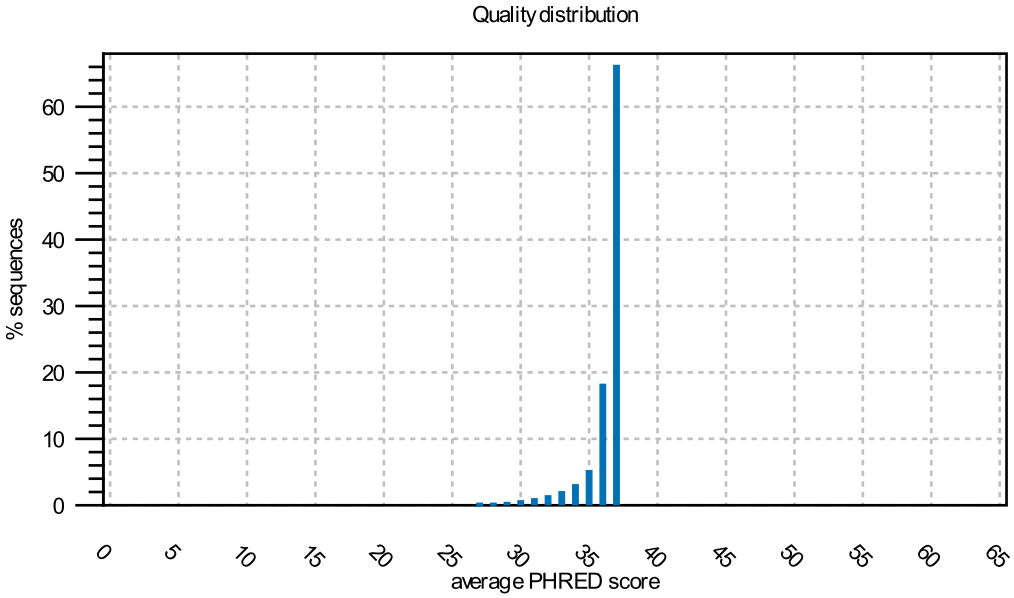
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

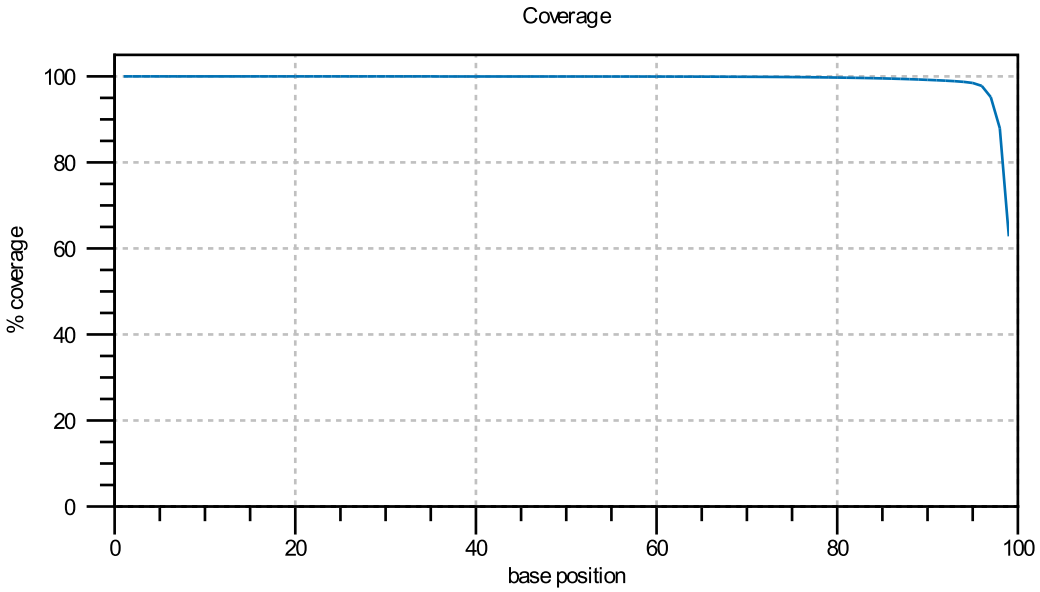
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

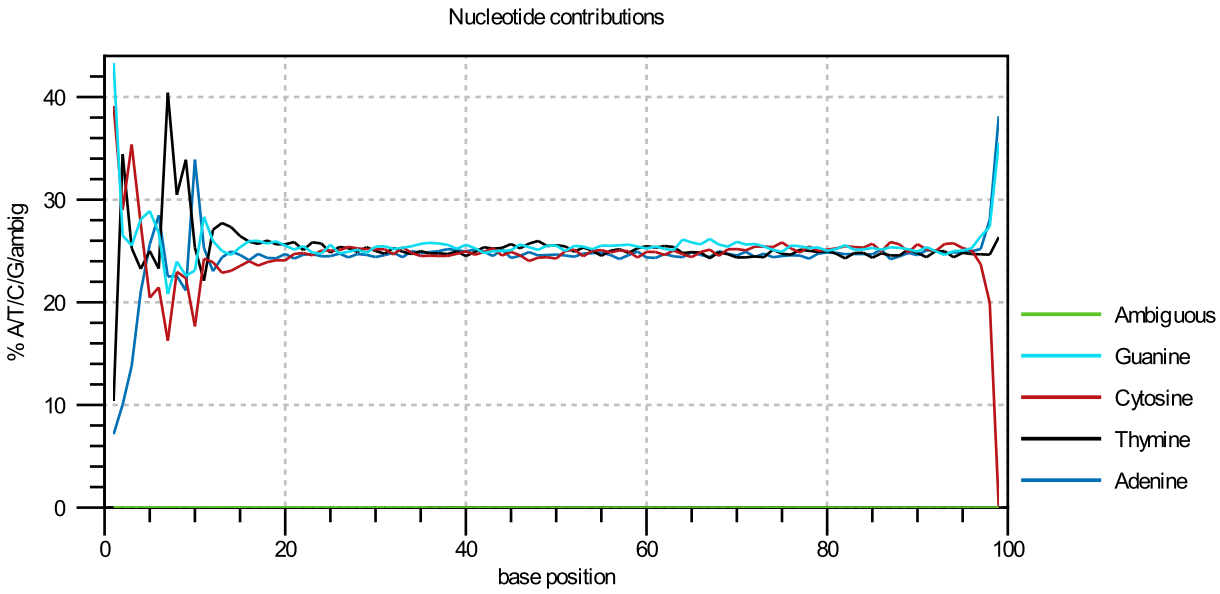
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

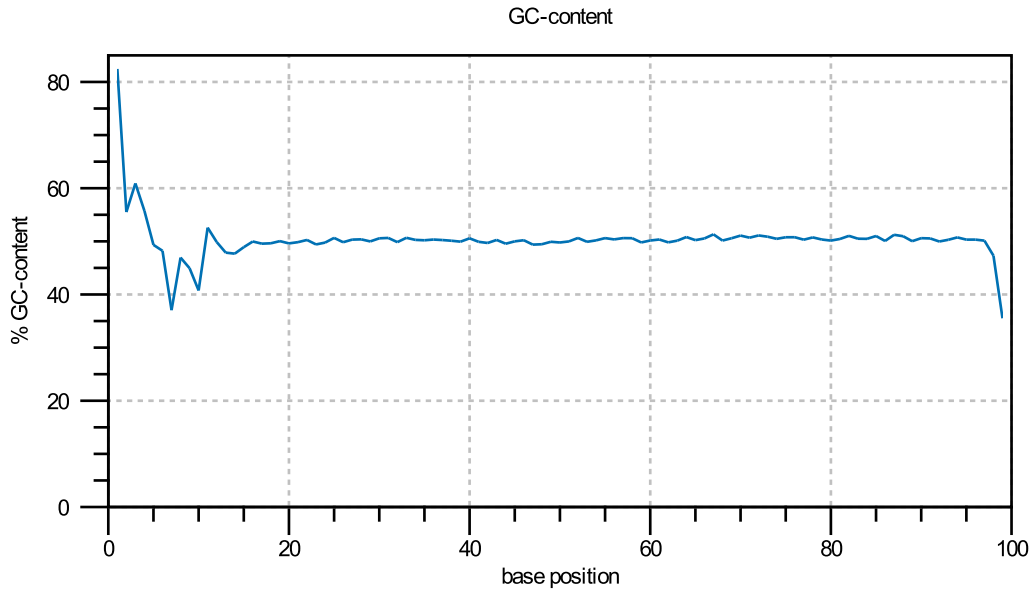
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

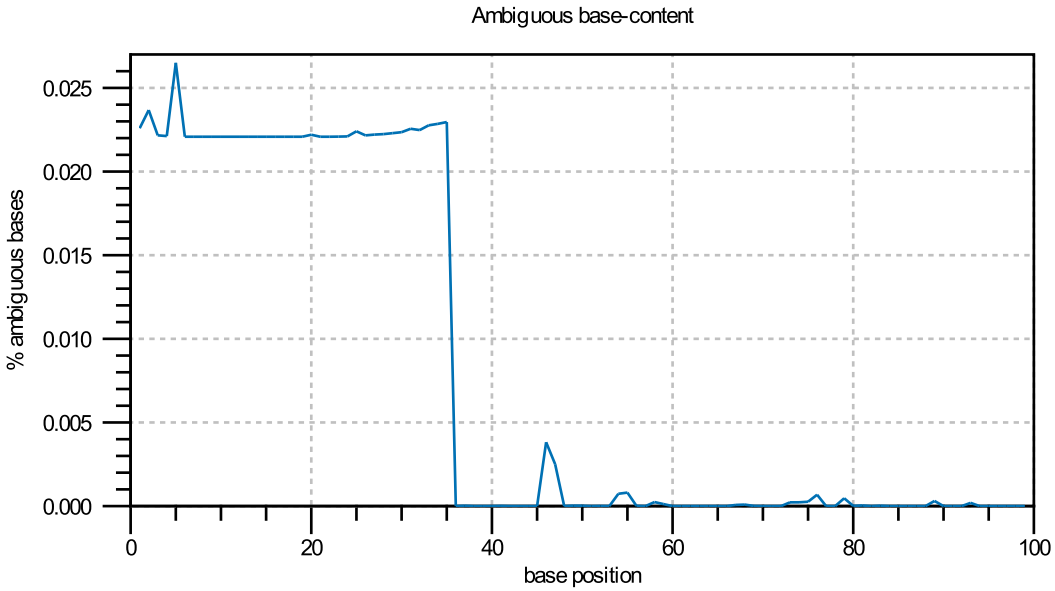
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



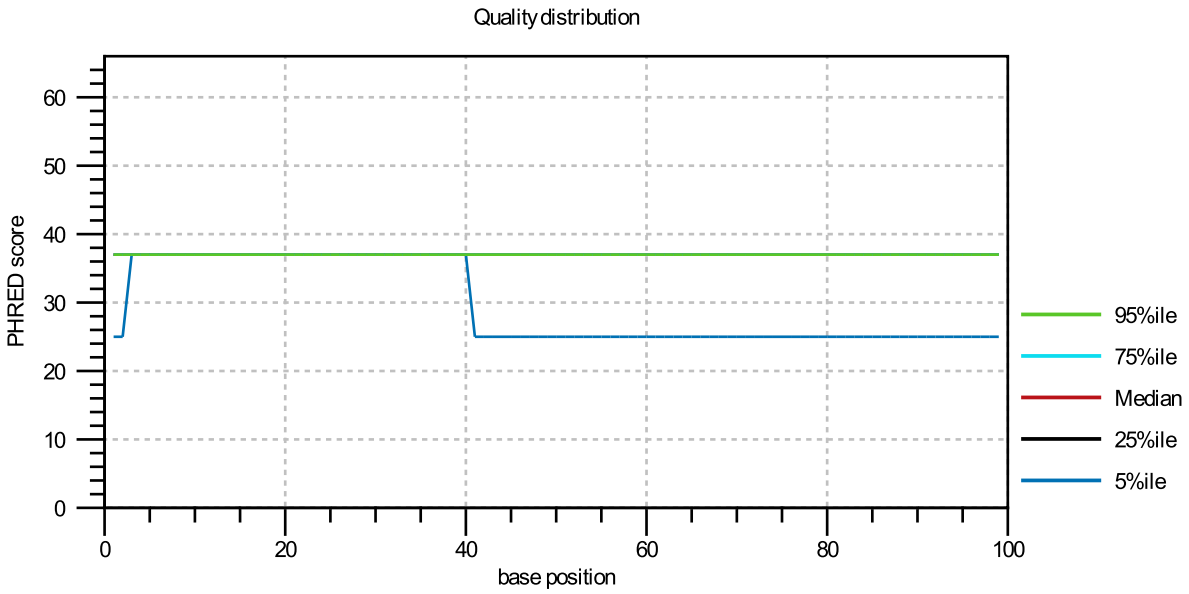
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

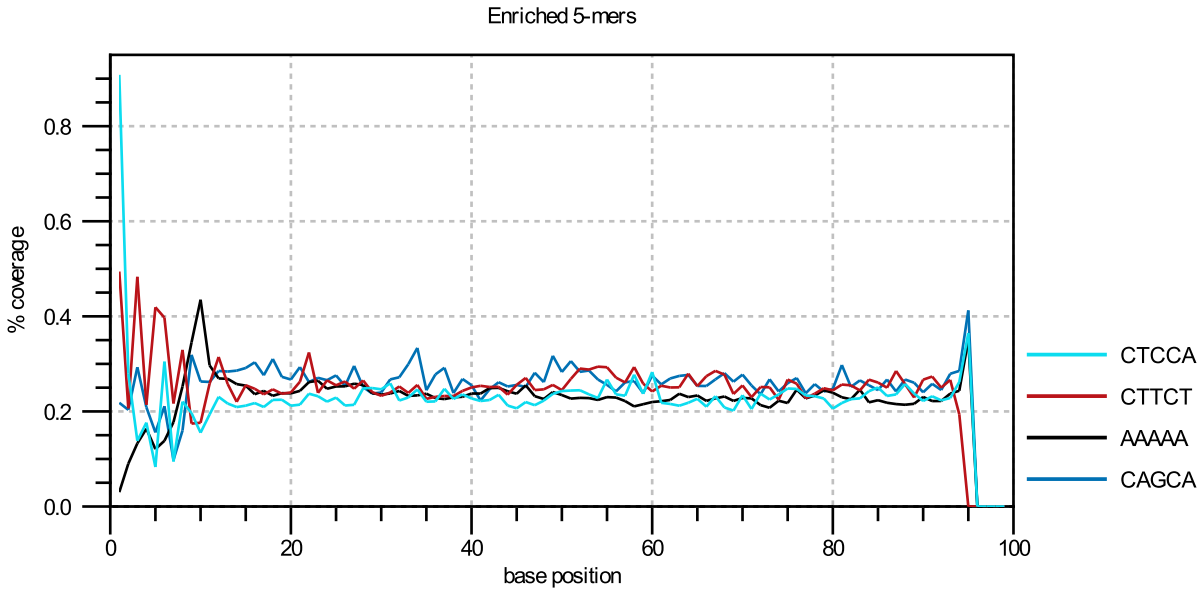
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

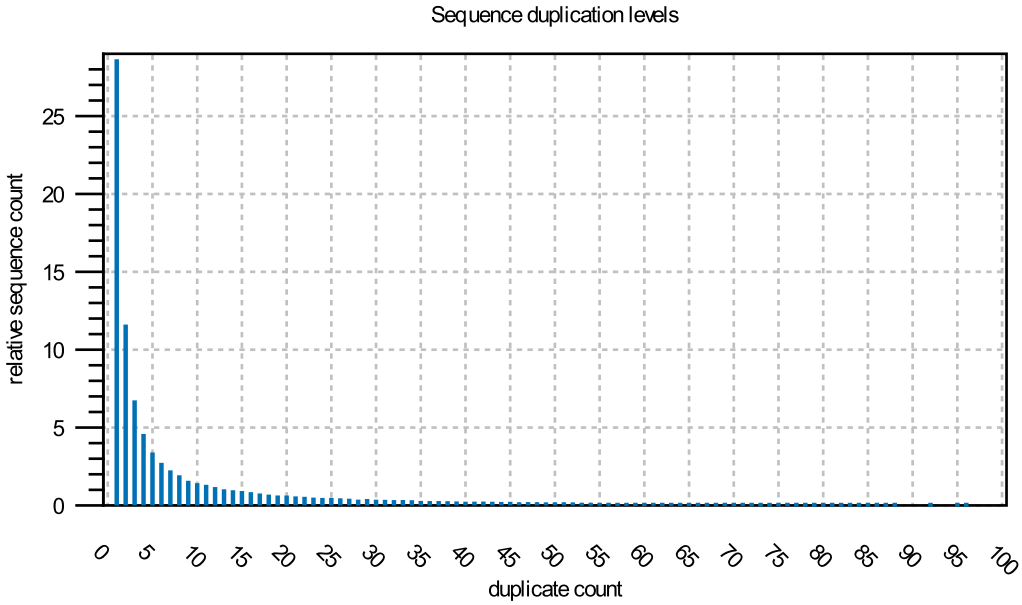
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

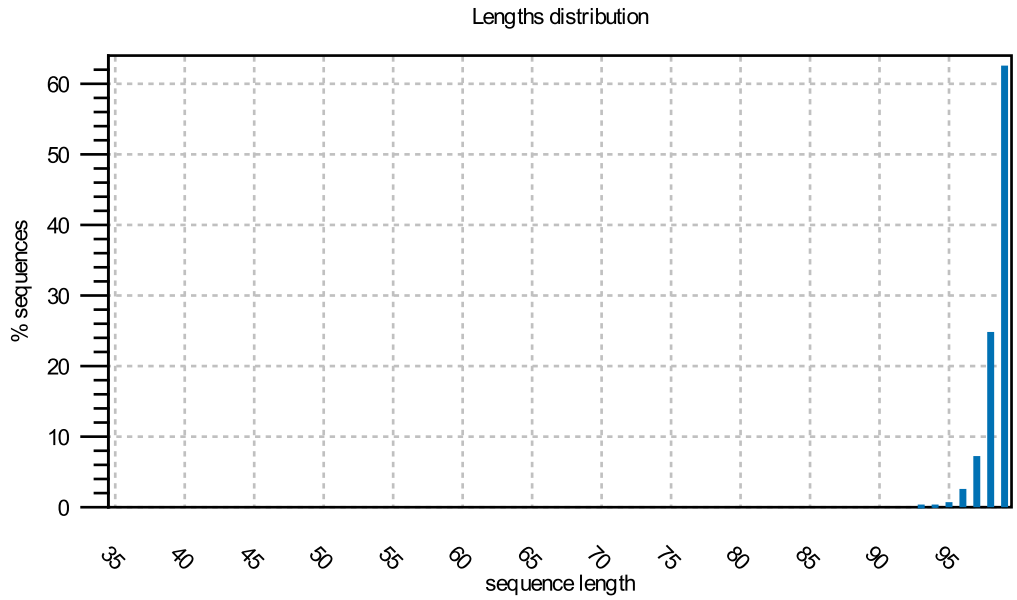
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 14:51:52 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
DL2_S3:	78,503,589 sequences
Total sequences in data set	78,503,589 sequences
Total nucleotides in data set	7,712,148,019 nucleotides

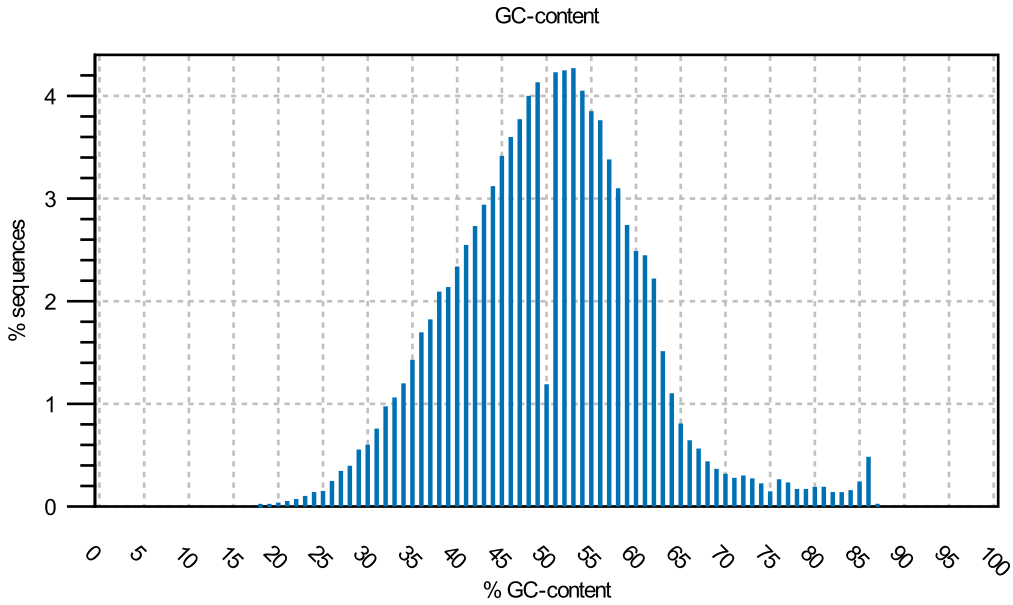
2. Per-sequence analysis

2.1 Lengths distribution



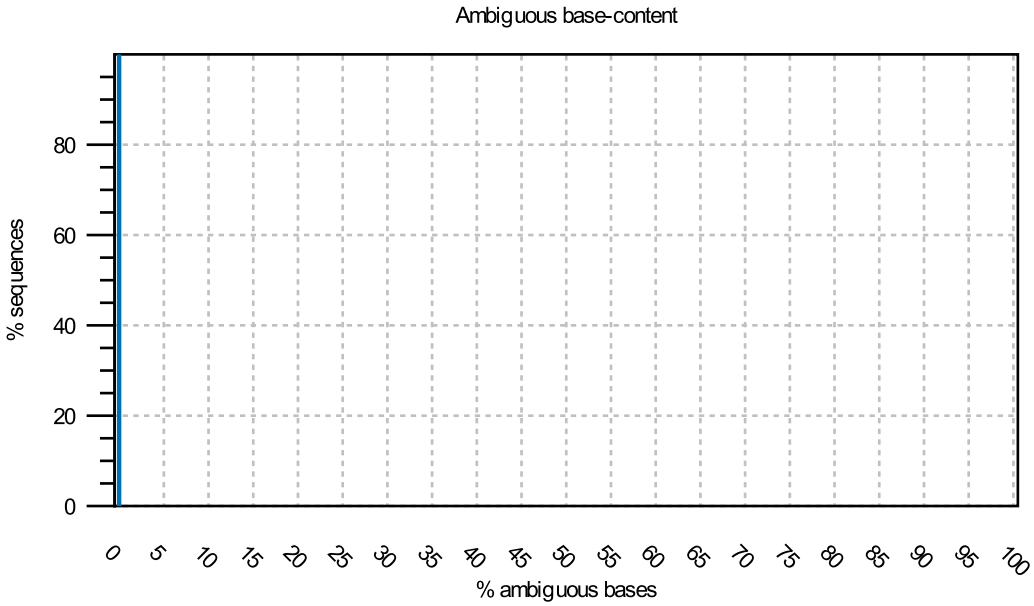
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



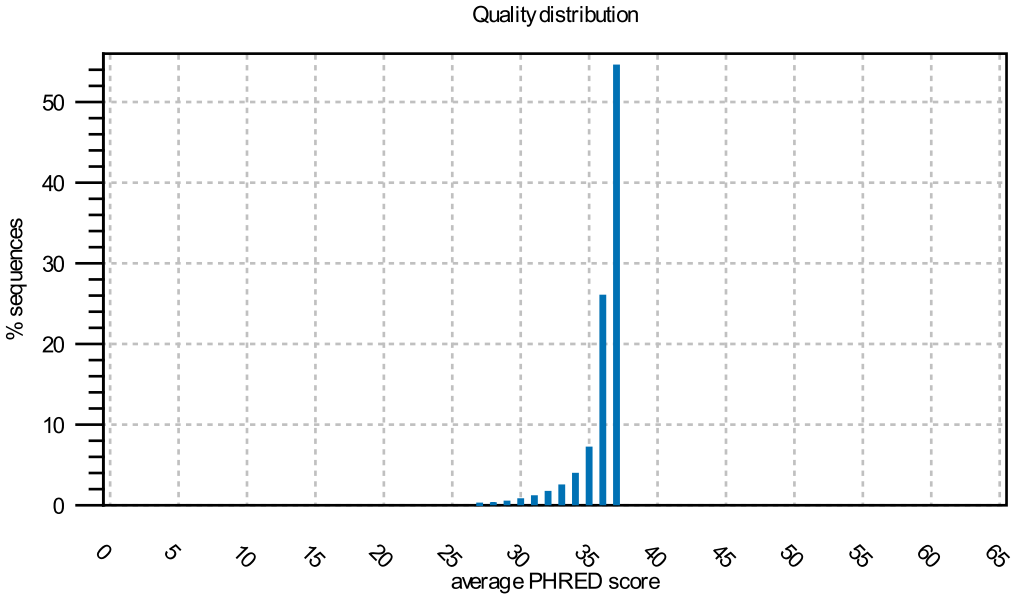
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

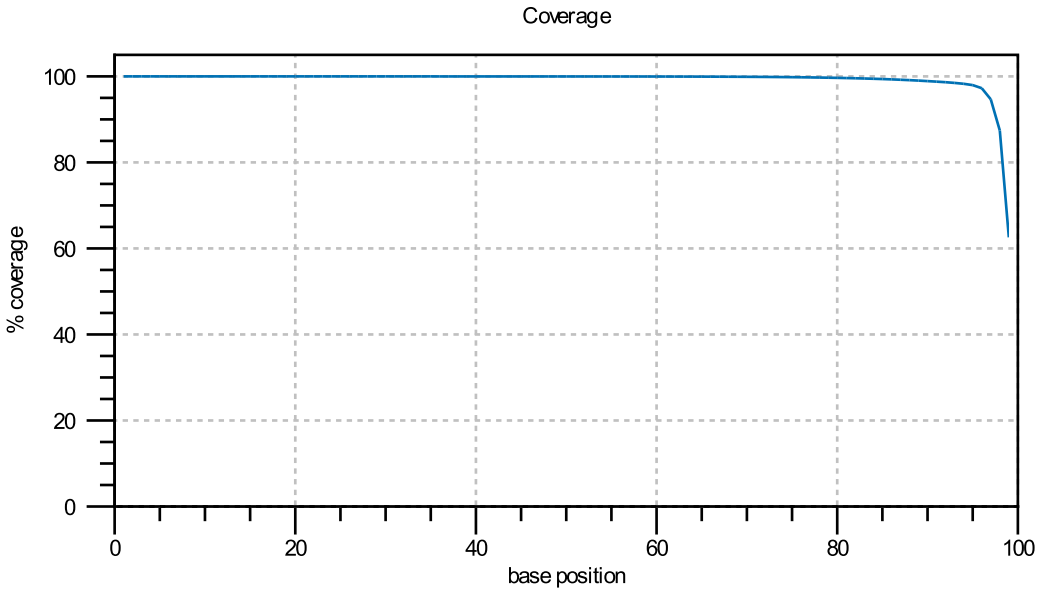
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

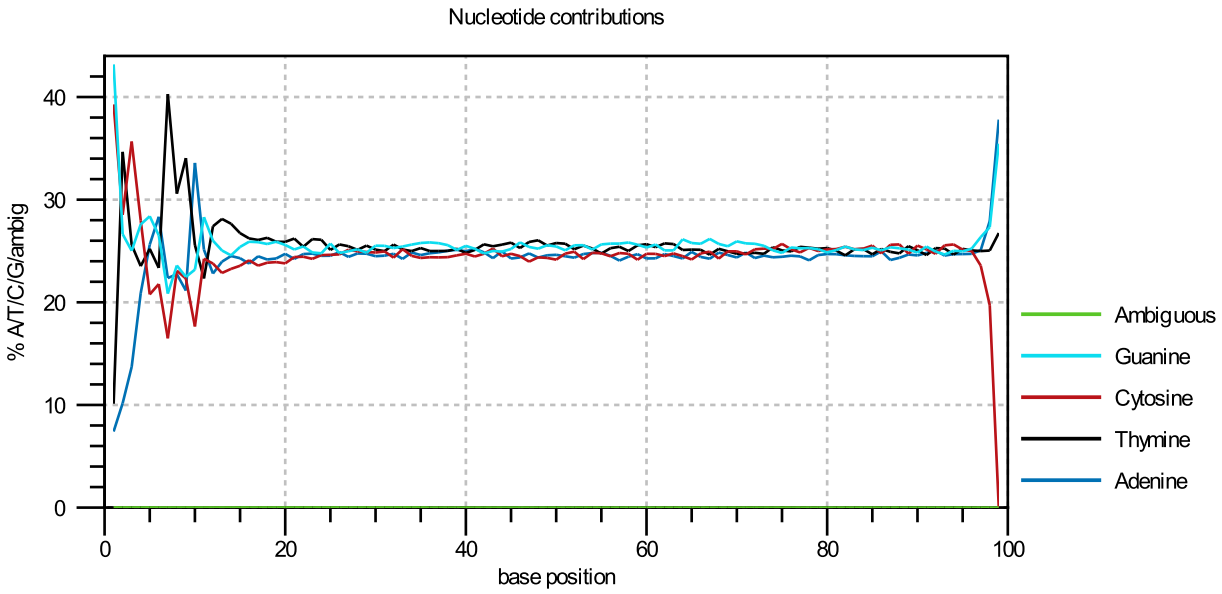
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

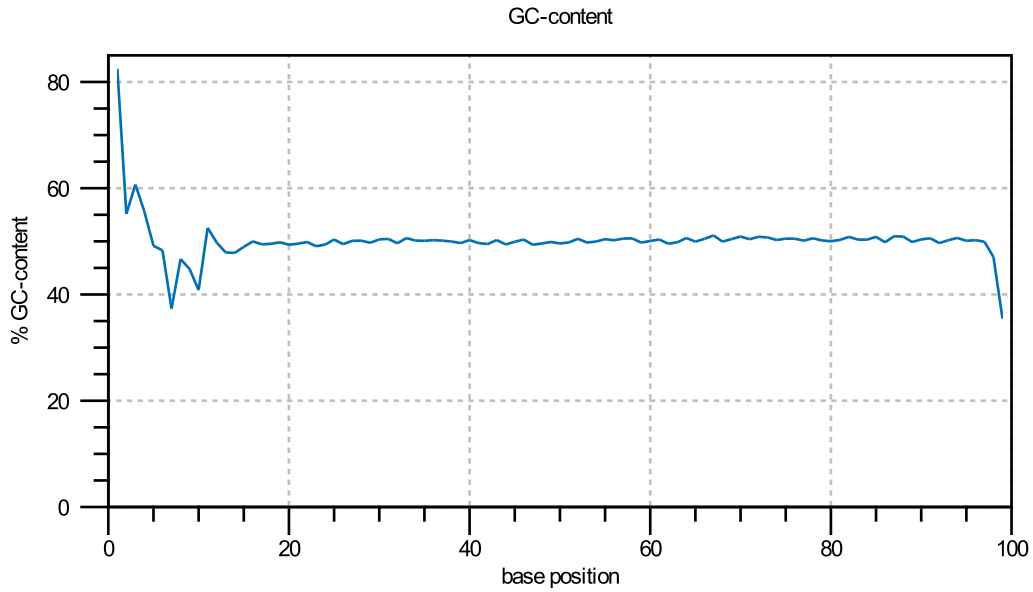
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

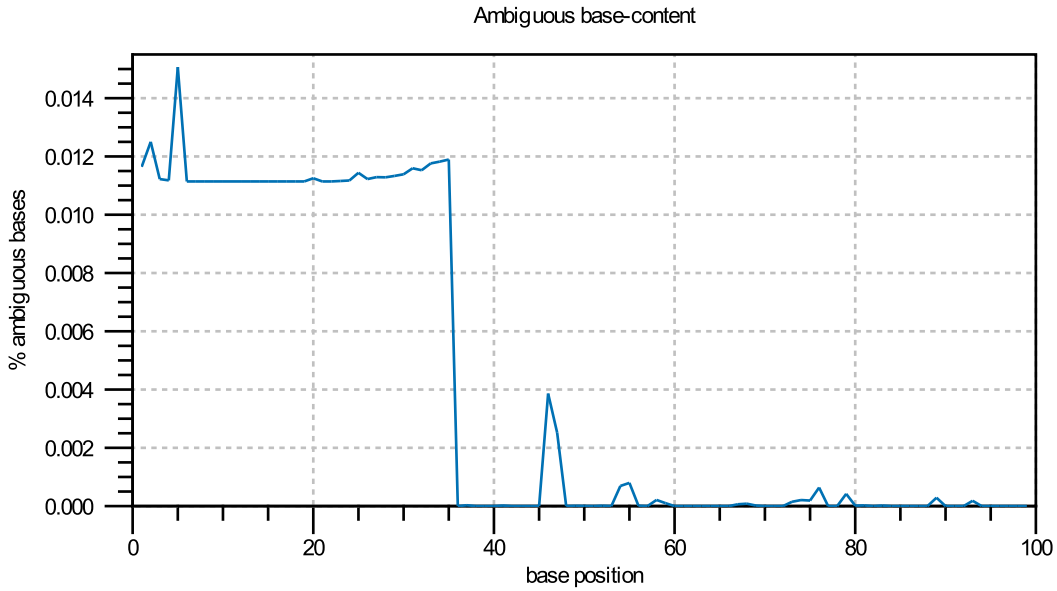
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



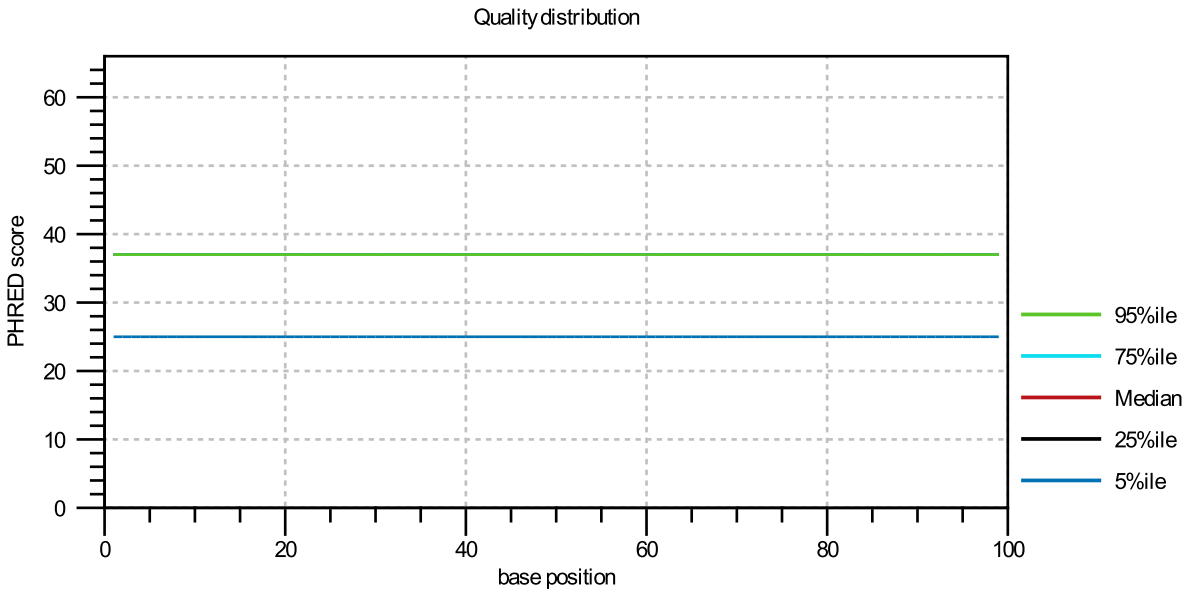
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

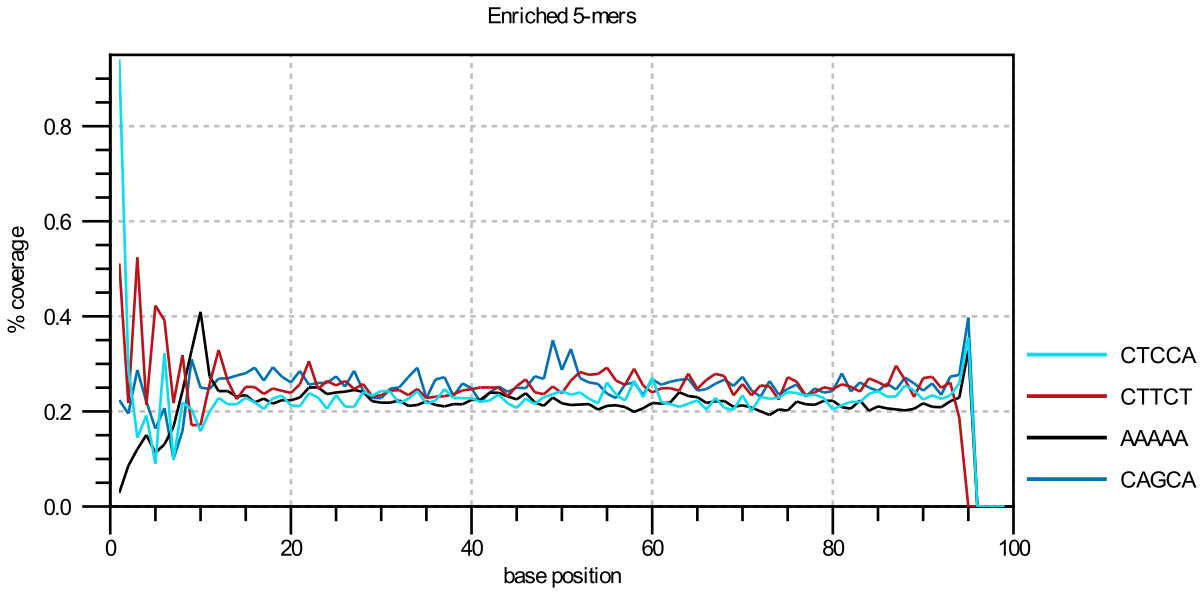
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

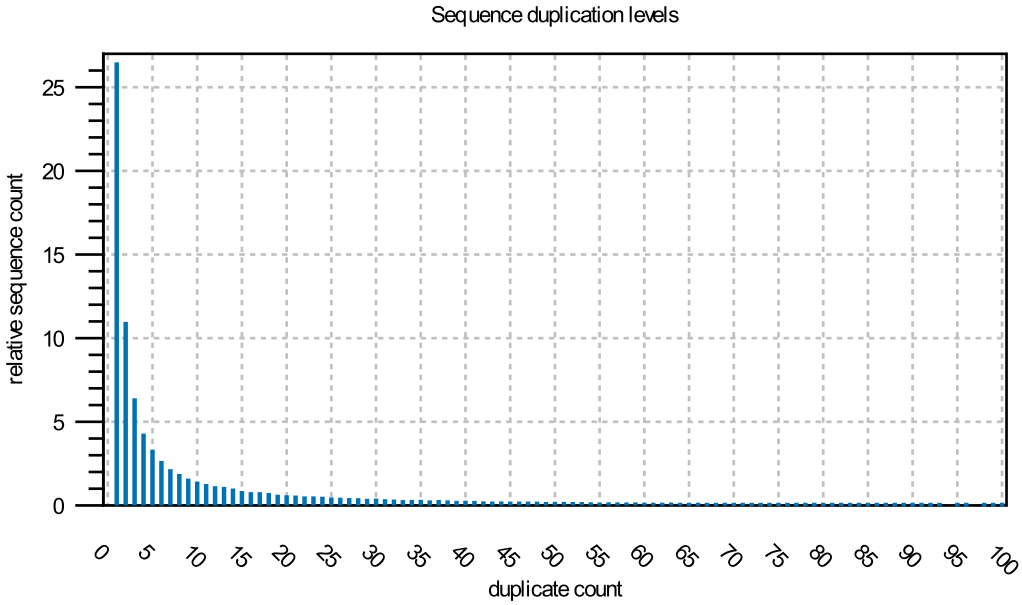
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

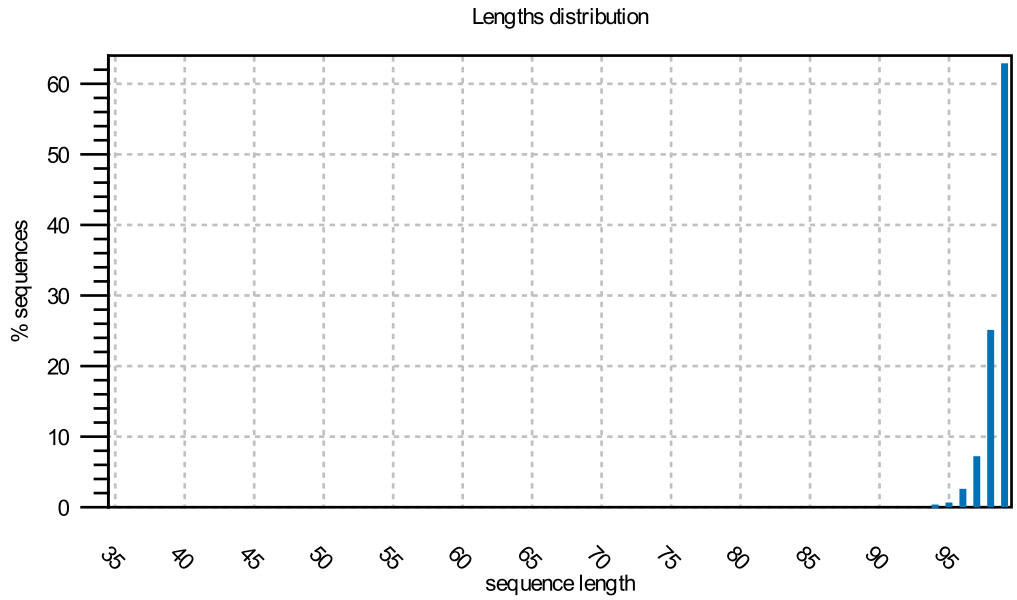
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 15:15:49 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
DL3_S2:	75,928,949 sequences
Total sequences in data set	75,928,949 sequences
Total nucleotides in data set	7,464,883,580 nucleotides

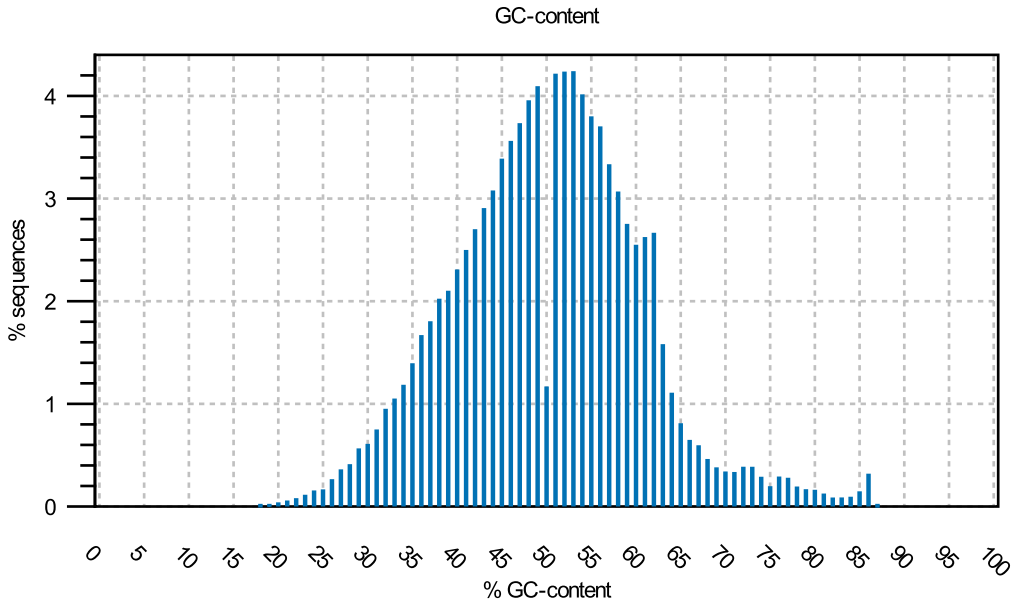
2. Per-sequence analysis

2.1 Lengths distribution



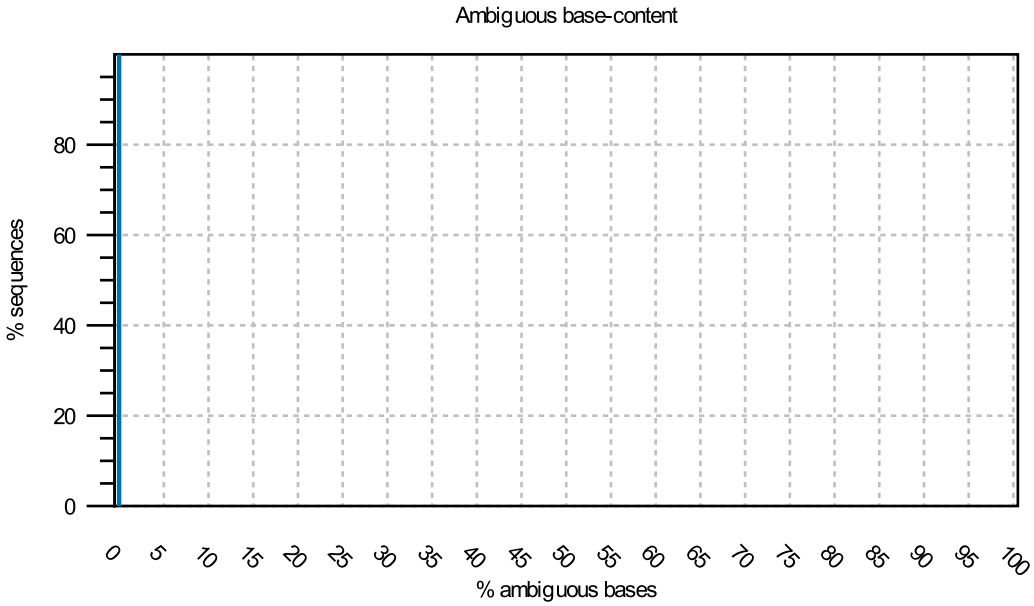
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



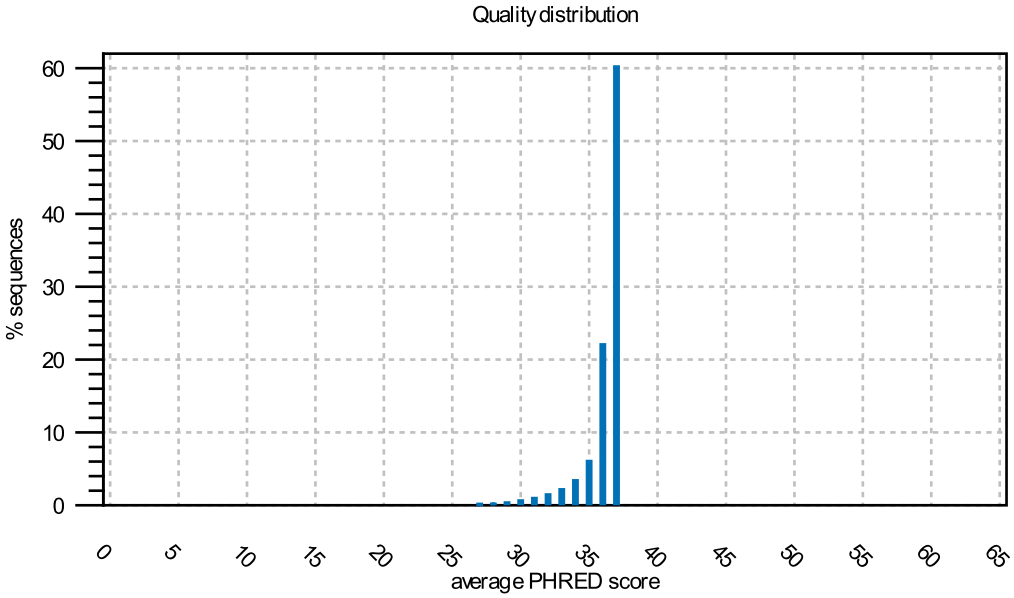
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

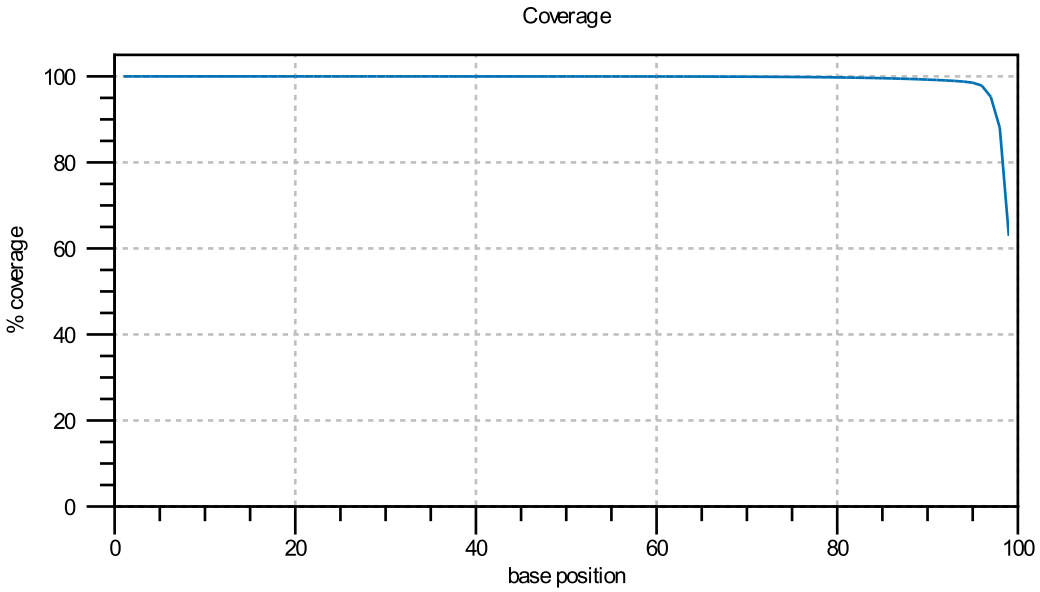
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

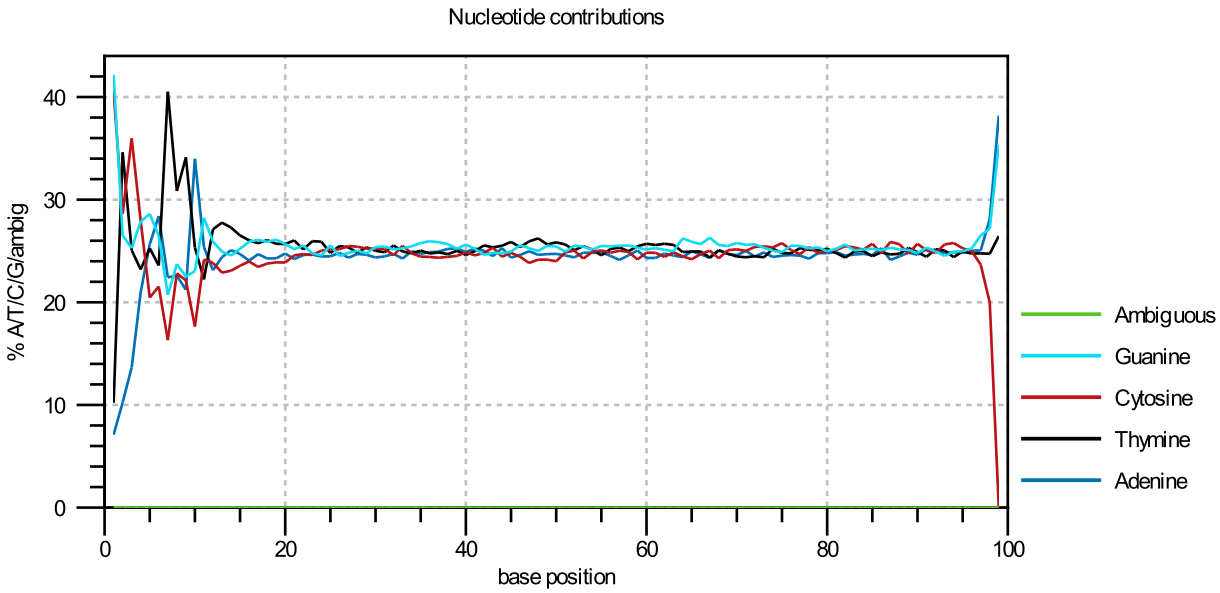
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

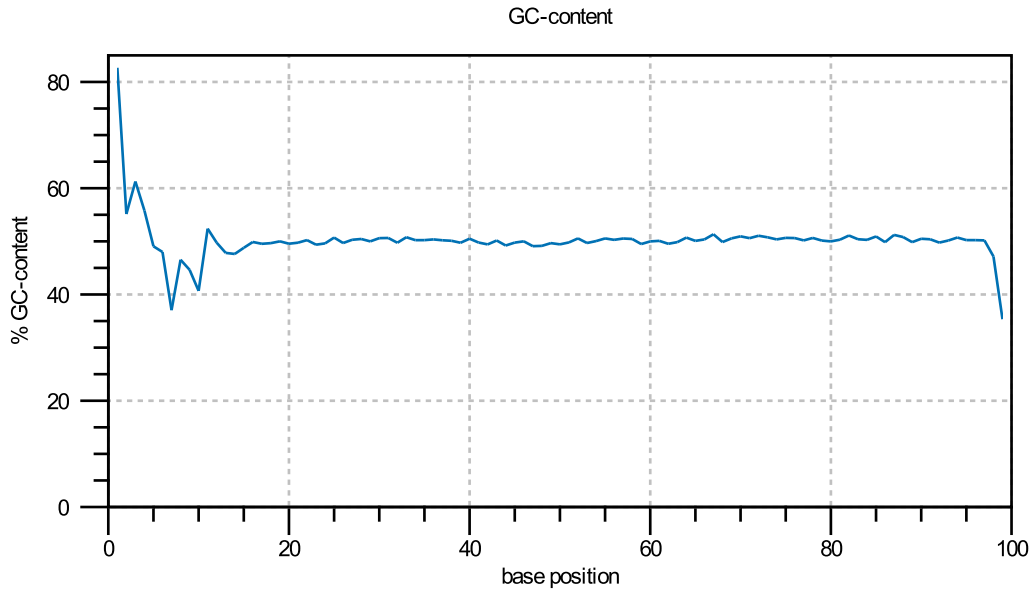
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

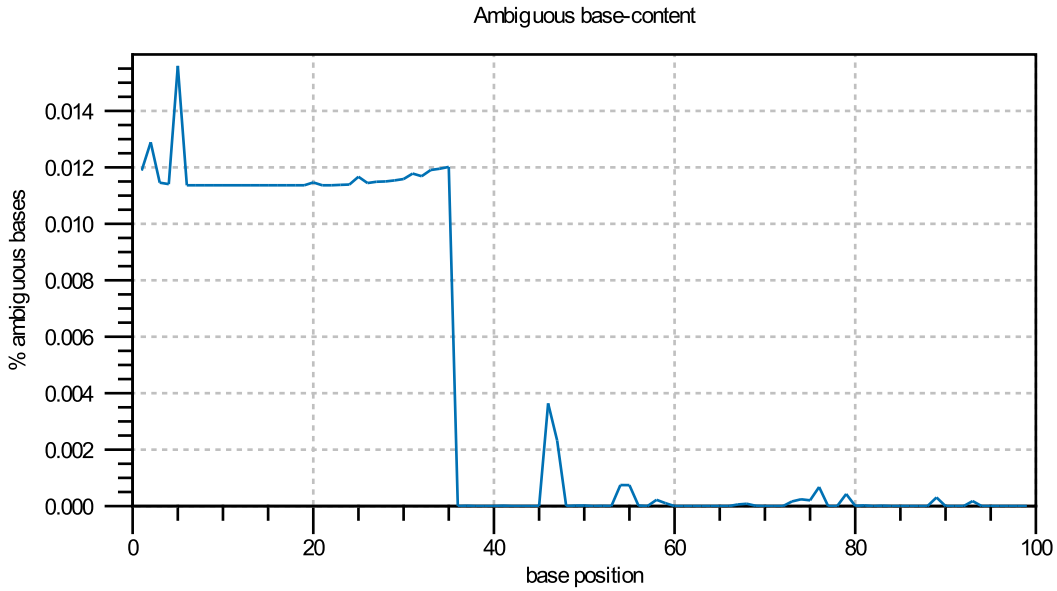
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



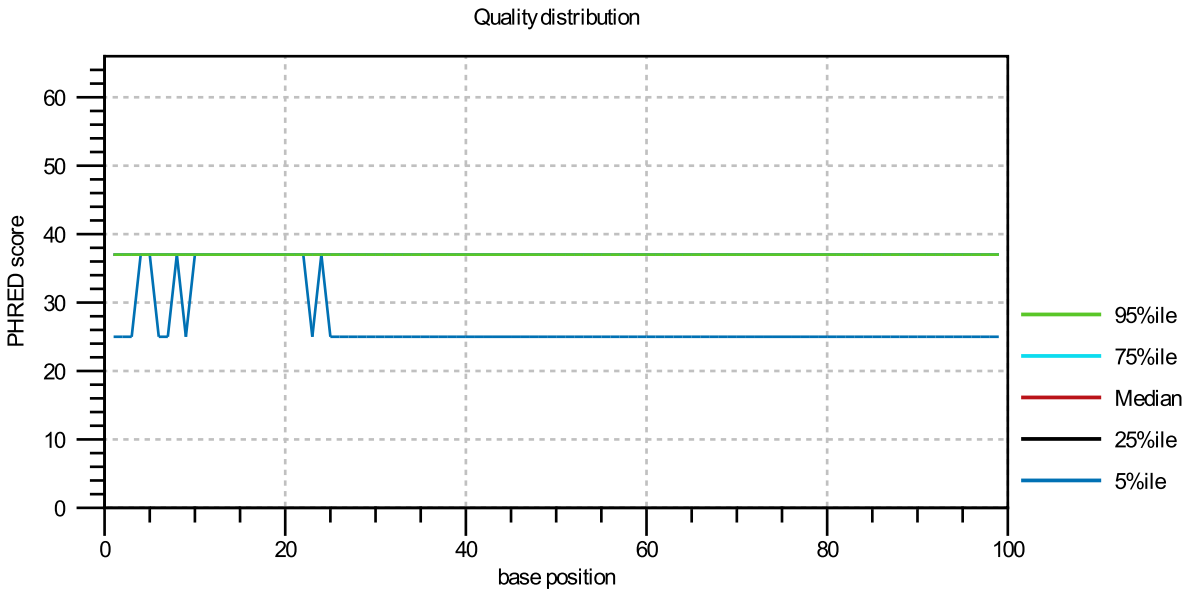
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

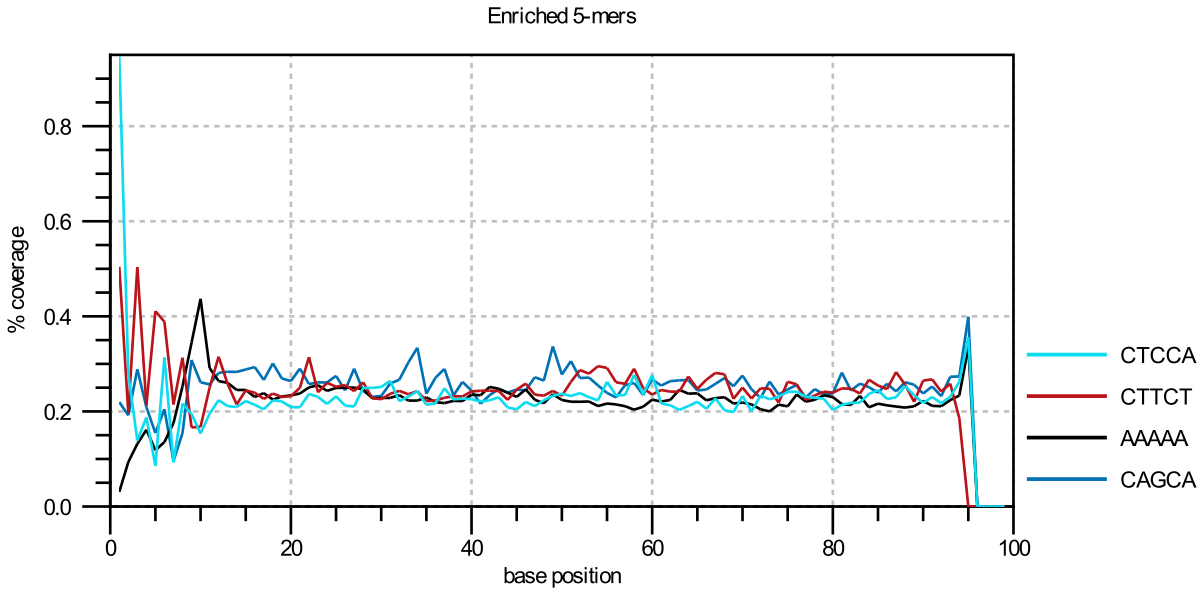
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

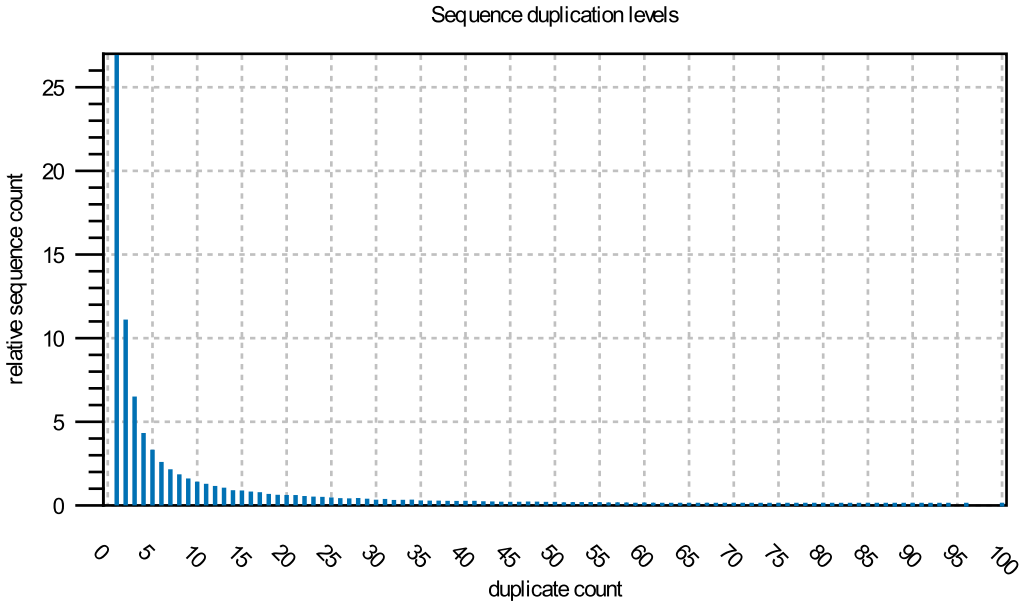
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

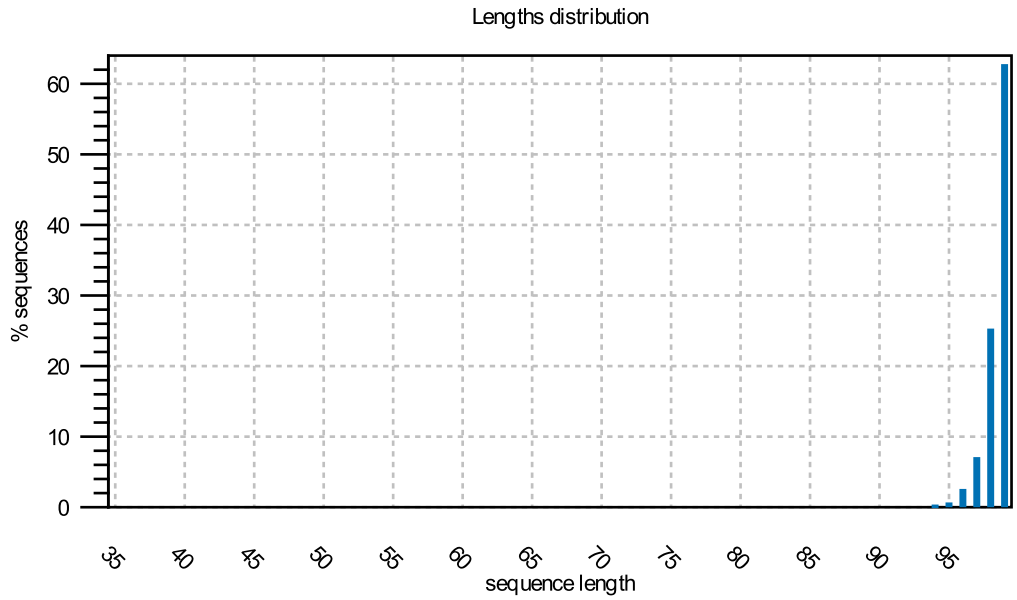
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 12:28:20 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
AAV1_S12:	73,054,598 sequences
Total sequences in data set	73,054,598 sequences
Total nucleotides in data set	7,181,818,285 nucleotides

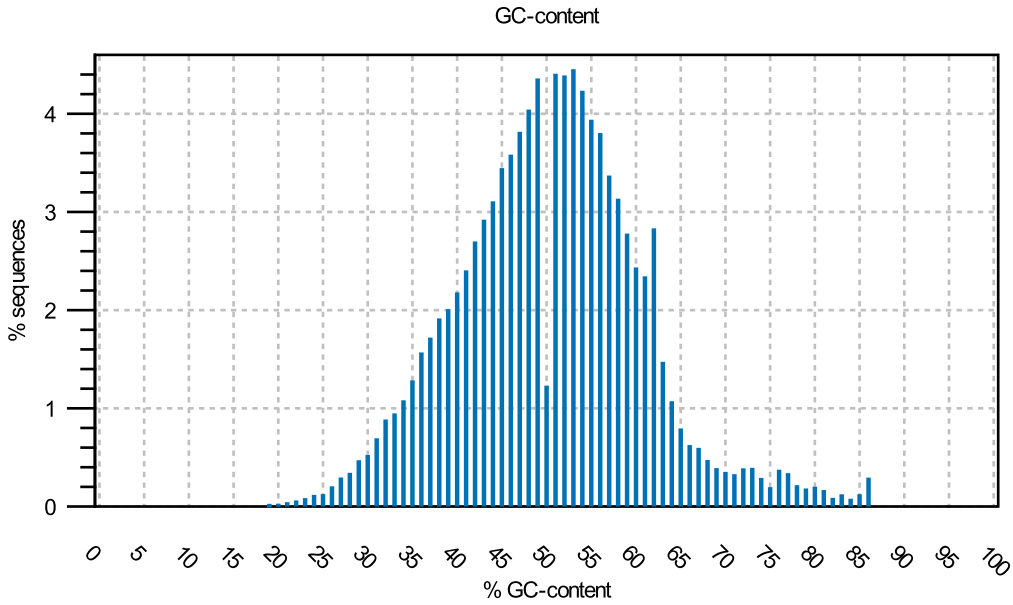
2. Per-sequence analysis

2.1 Lengths distribution



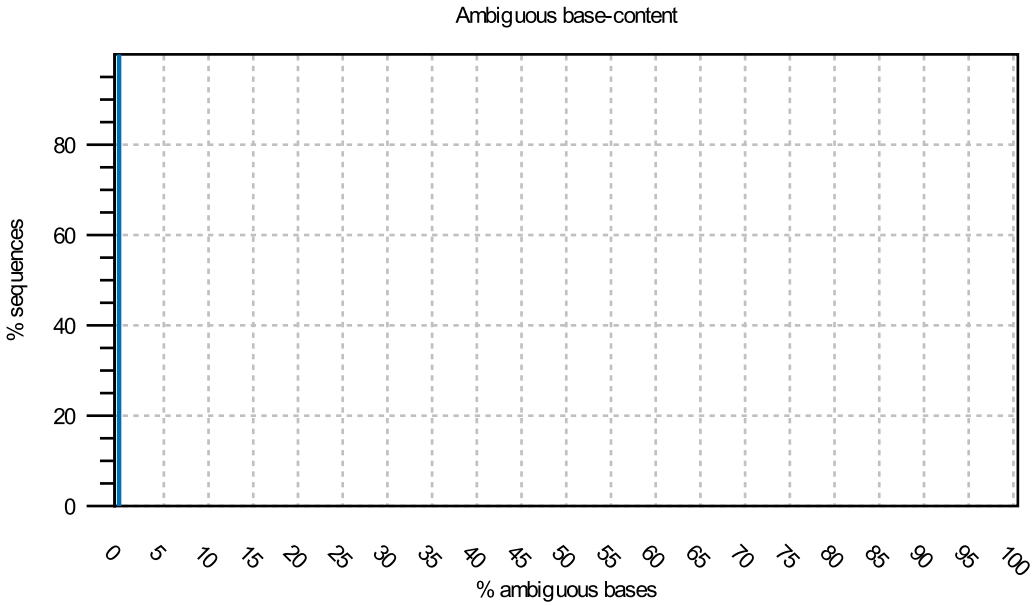
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



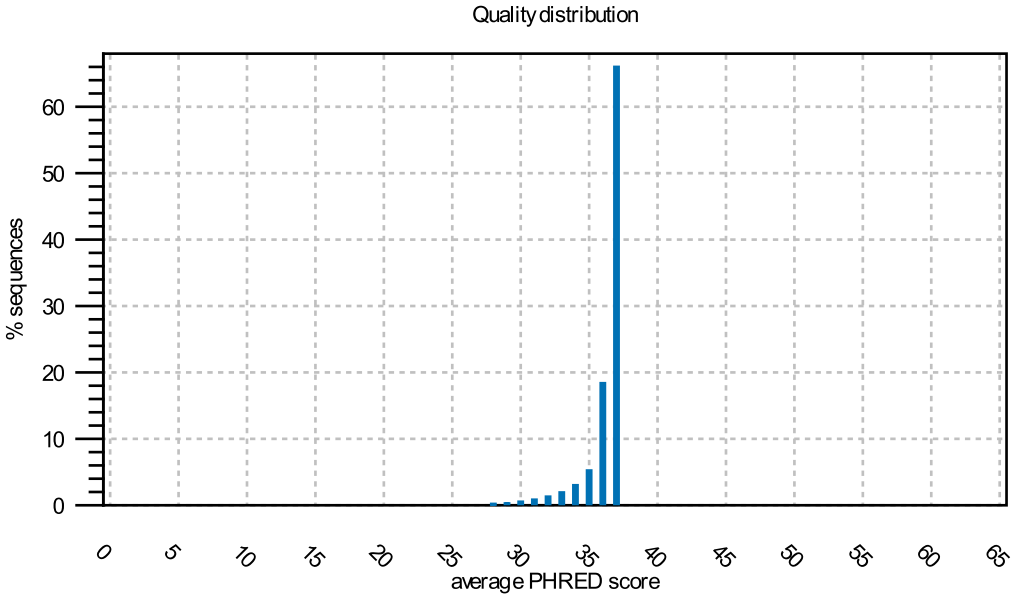
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

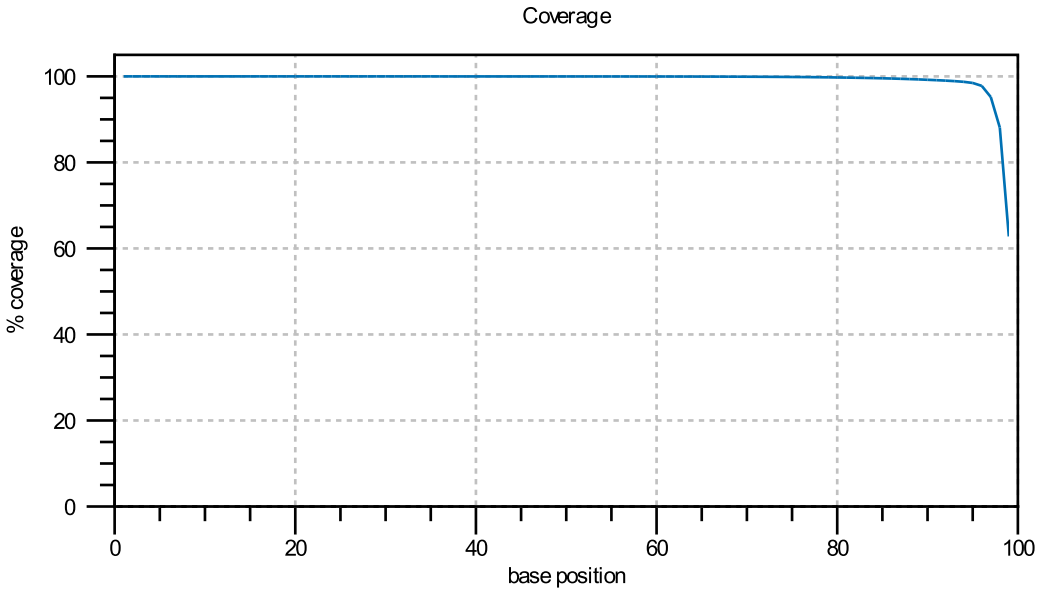
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

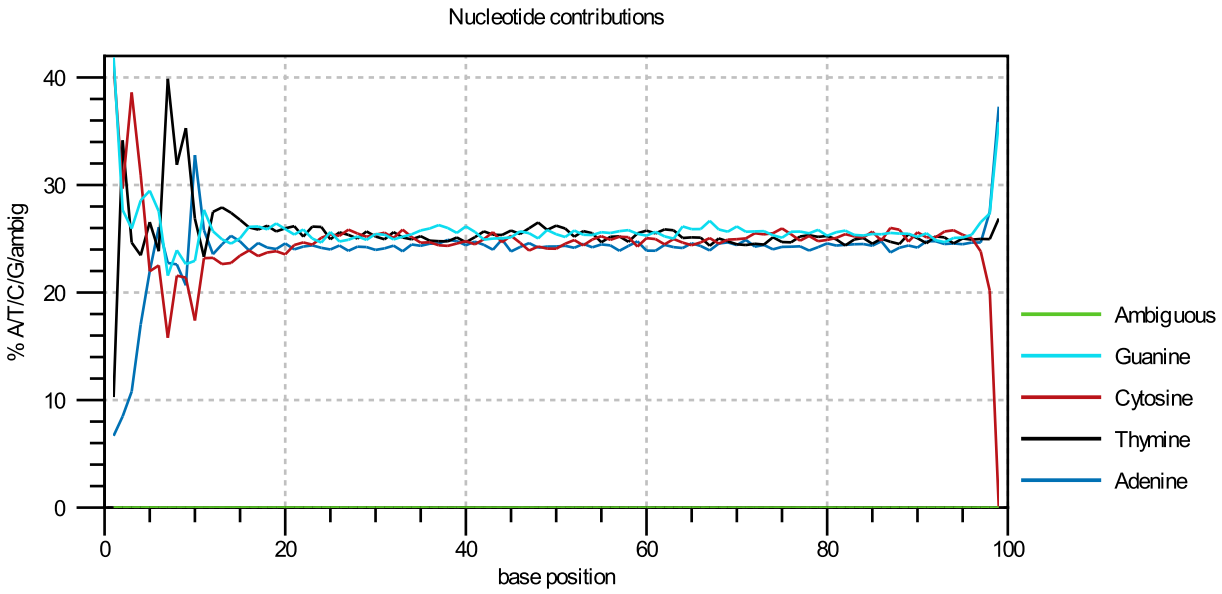
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

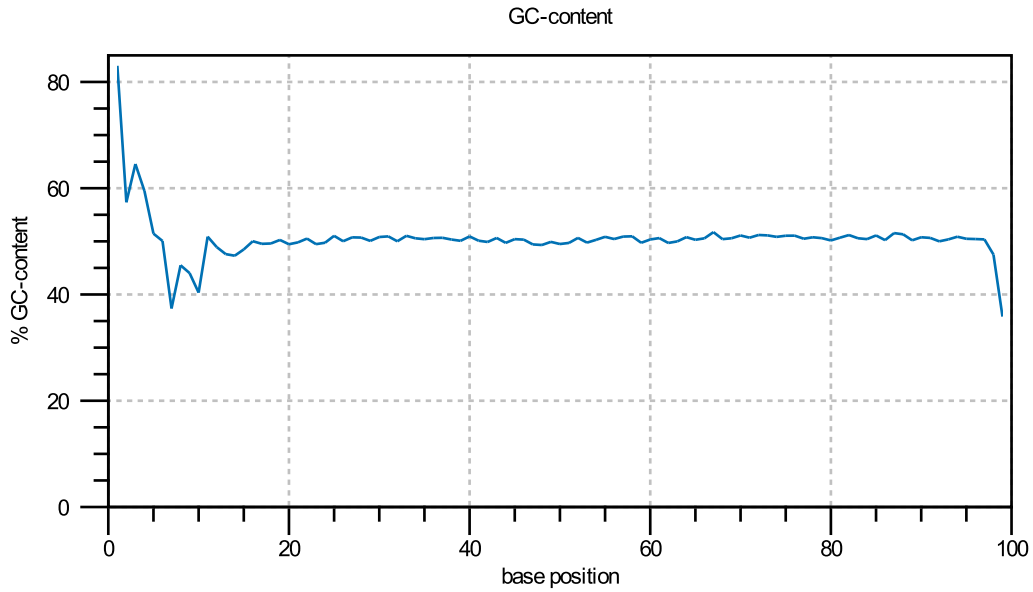
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

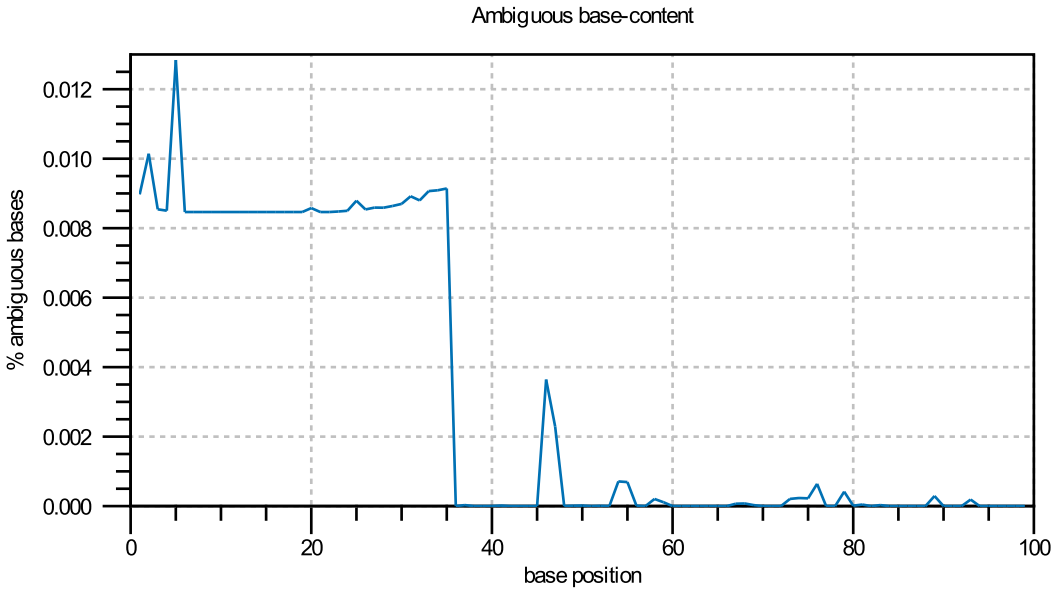
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



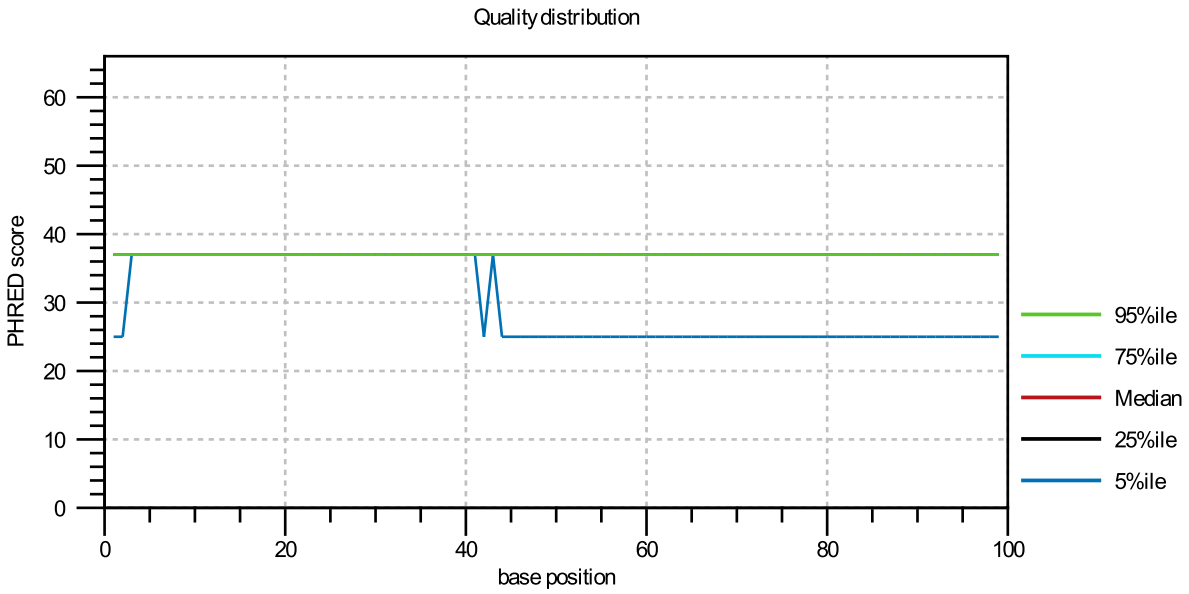
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

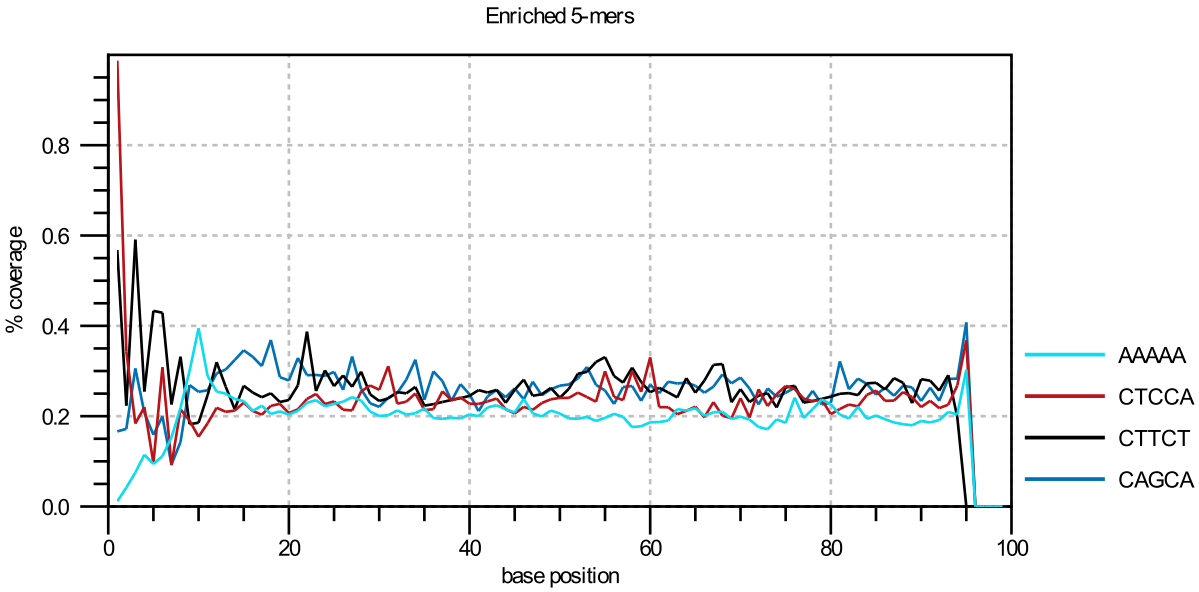
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

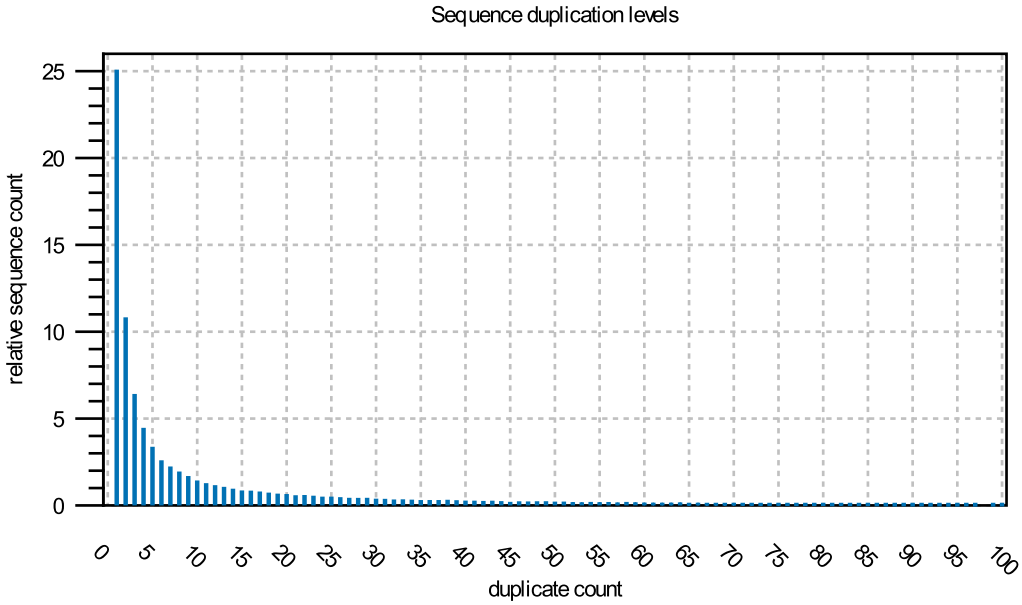
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

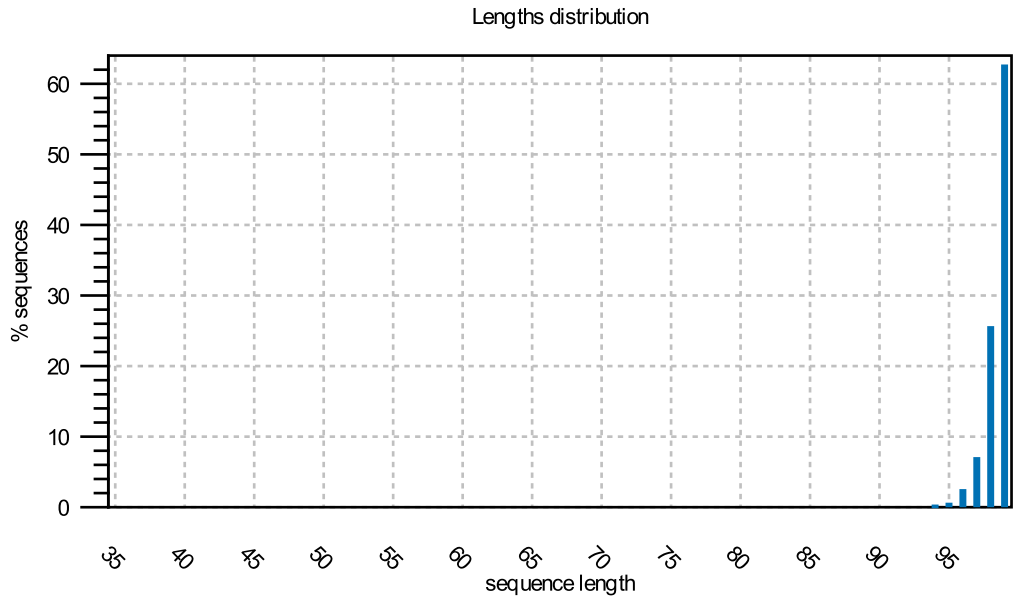
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 12:50:25 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
AAV2_S9:	88,511,361 sequences
Total sequences in data set	88,511,361 sequences
Total nucleotides in data set	8,704,068,742 nucleotides

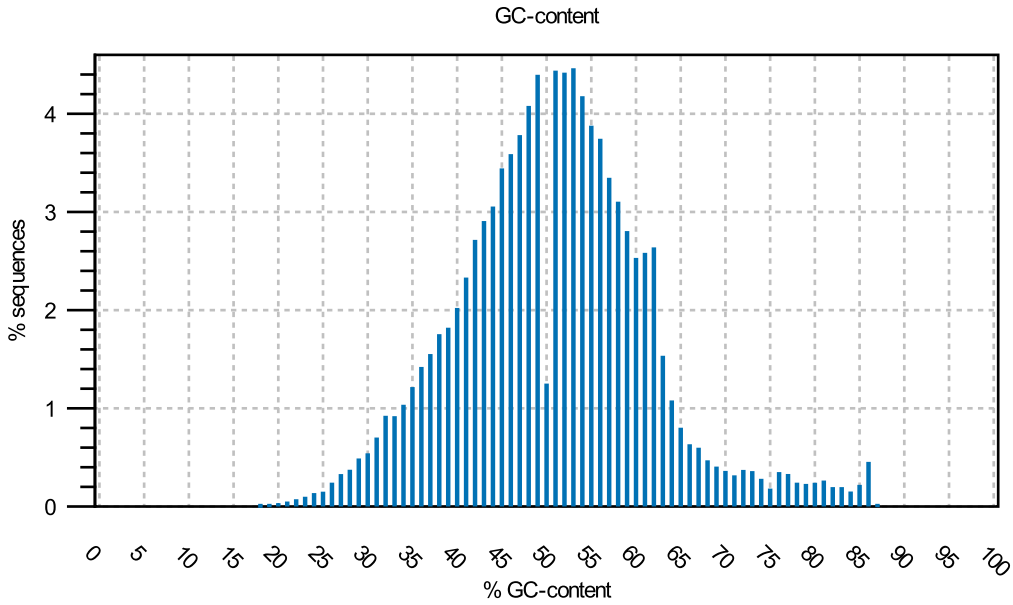
2. Per-sequence analysis

2.1 Lengths distribution



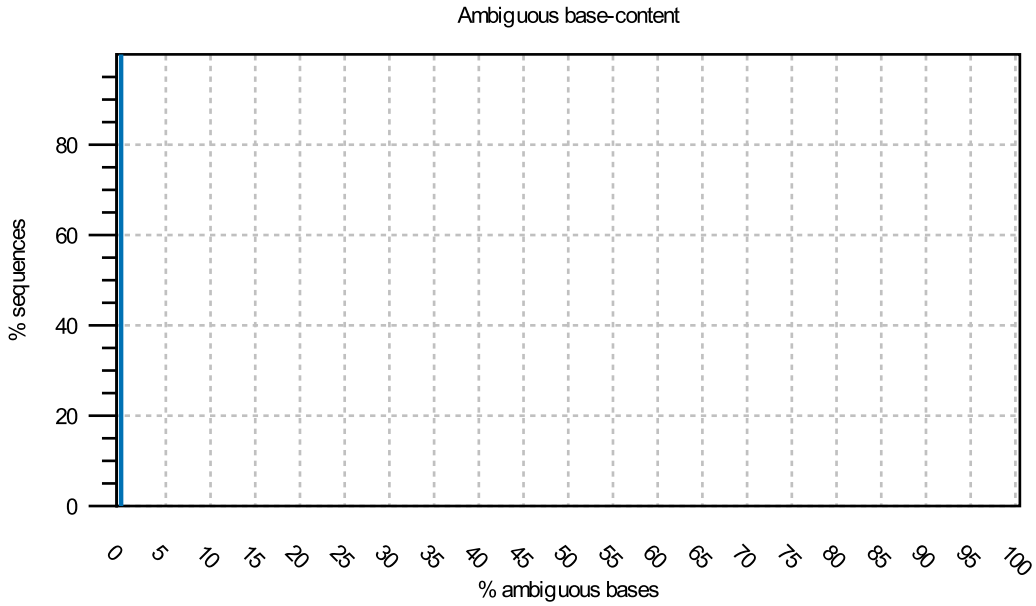
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



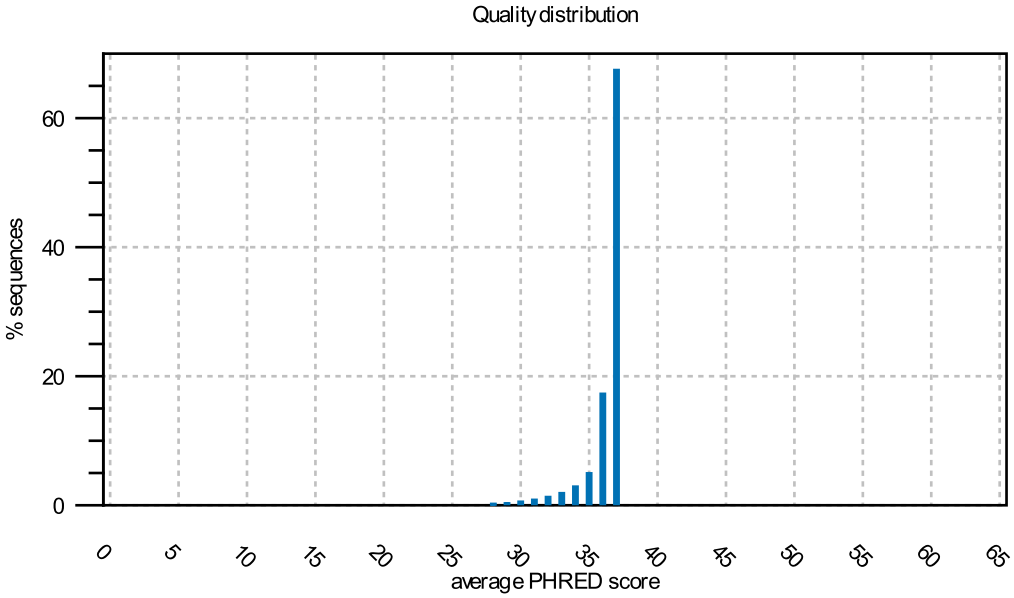
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

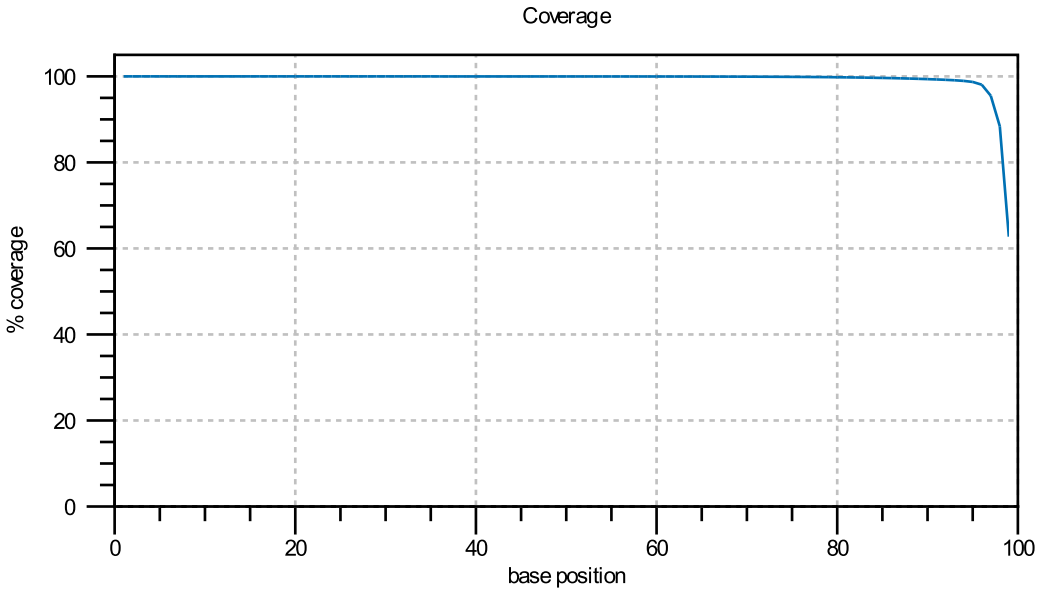
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

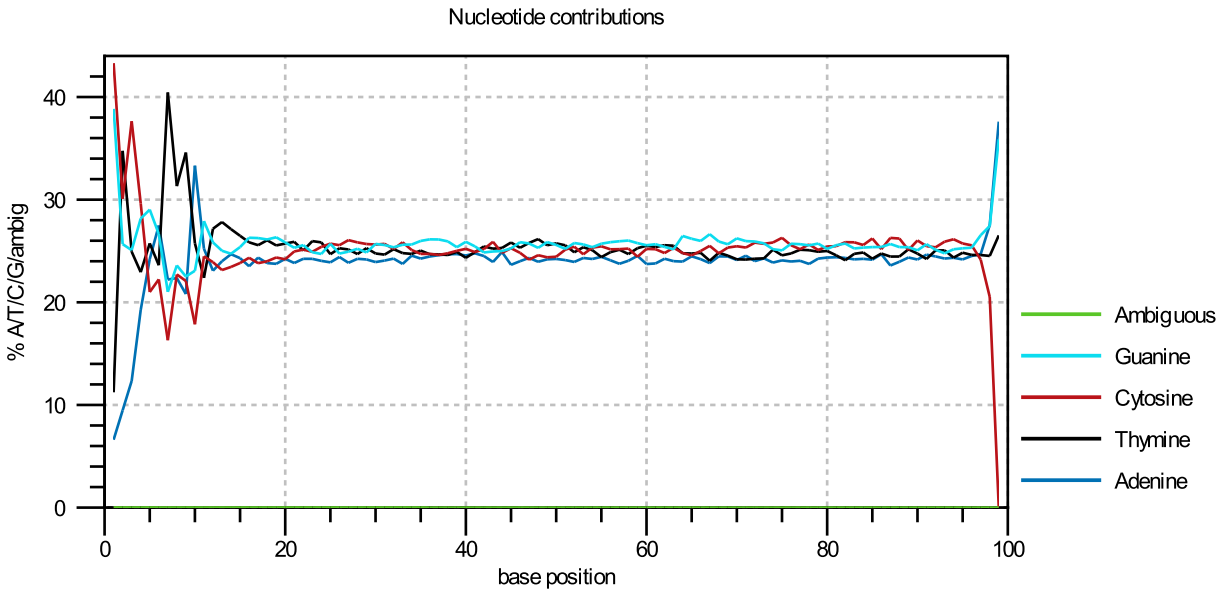
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

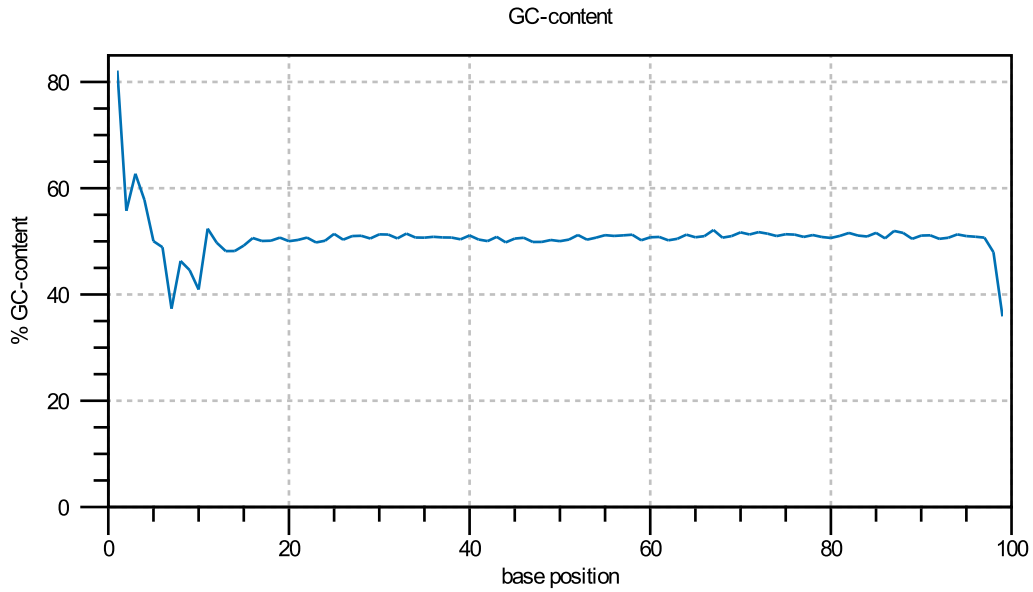
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

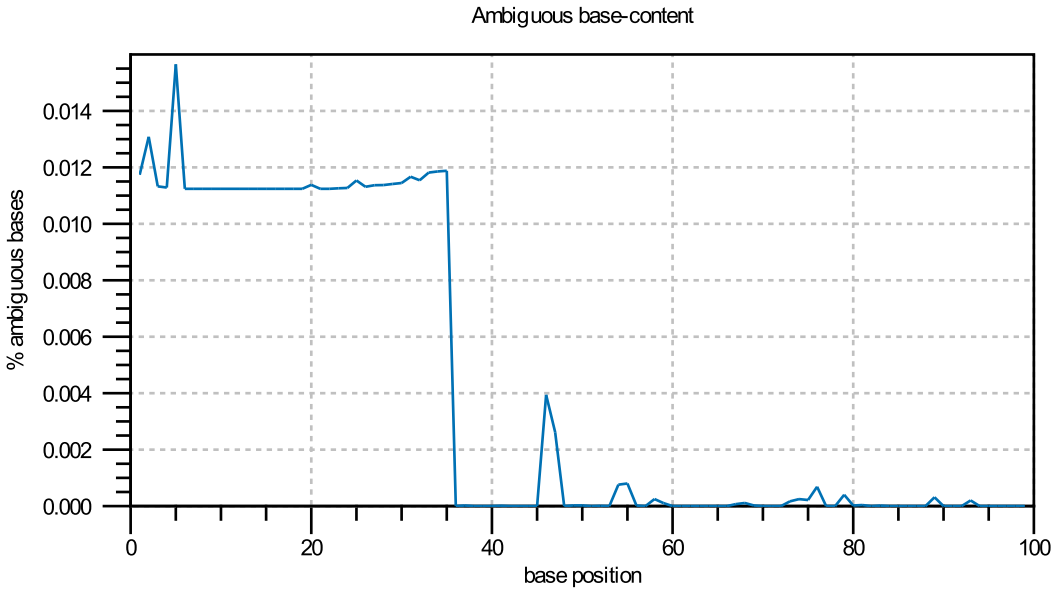
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



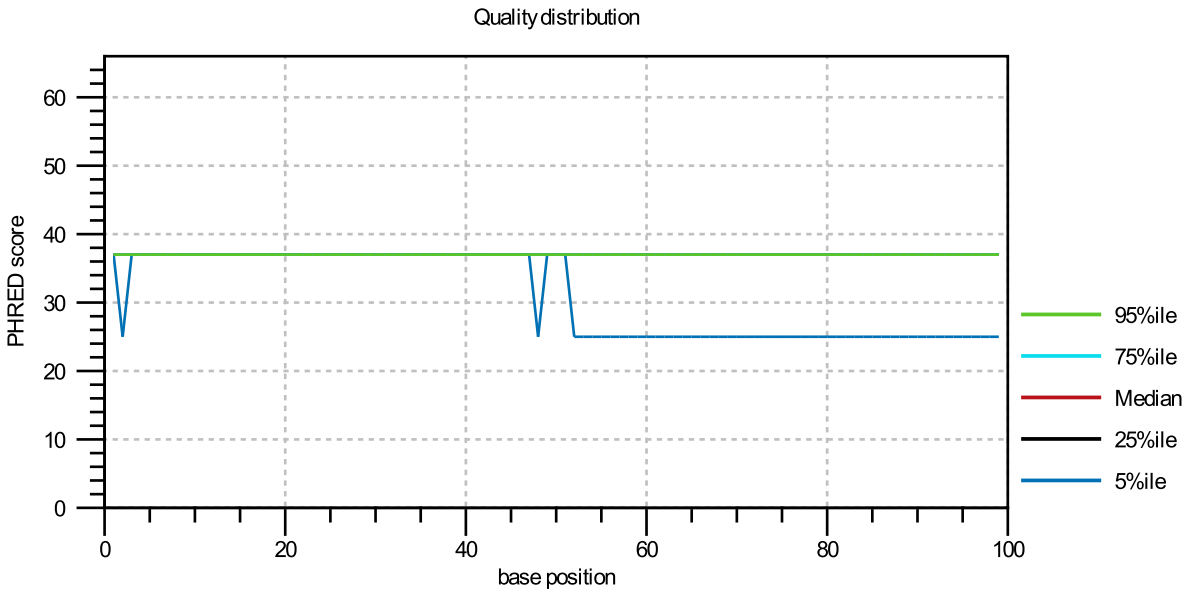
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

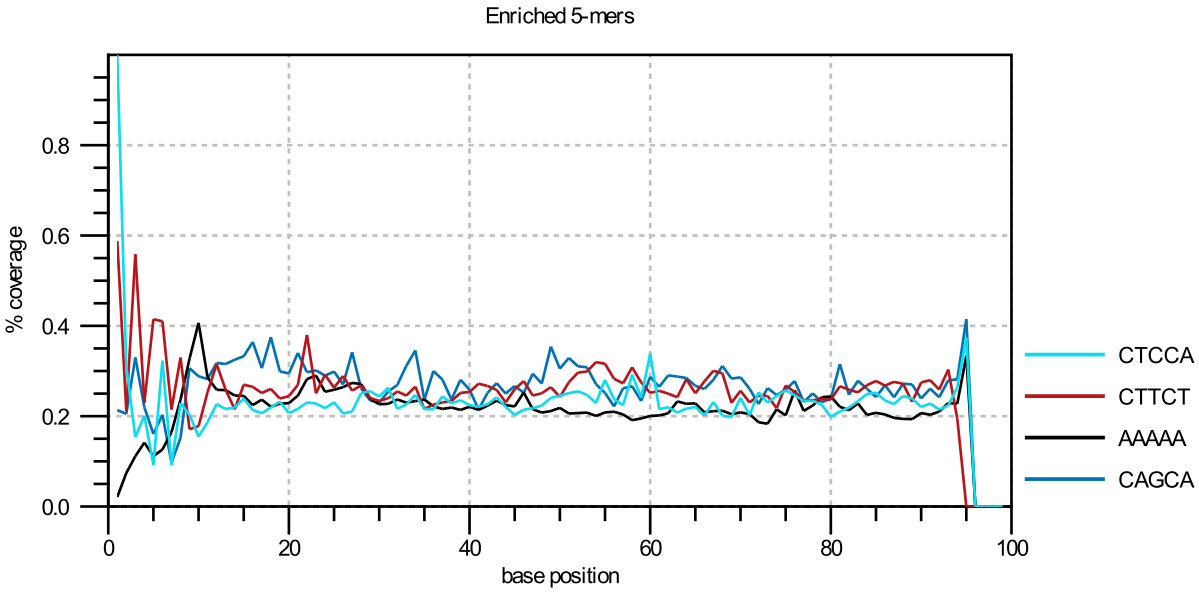
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

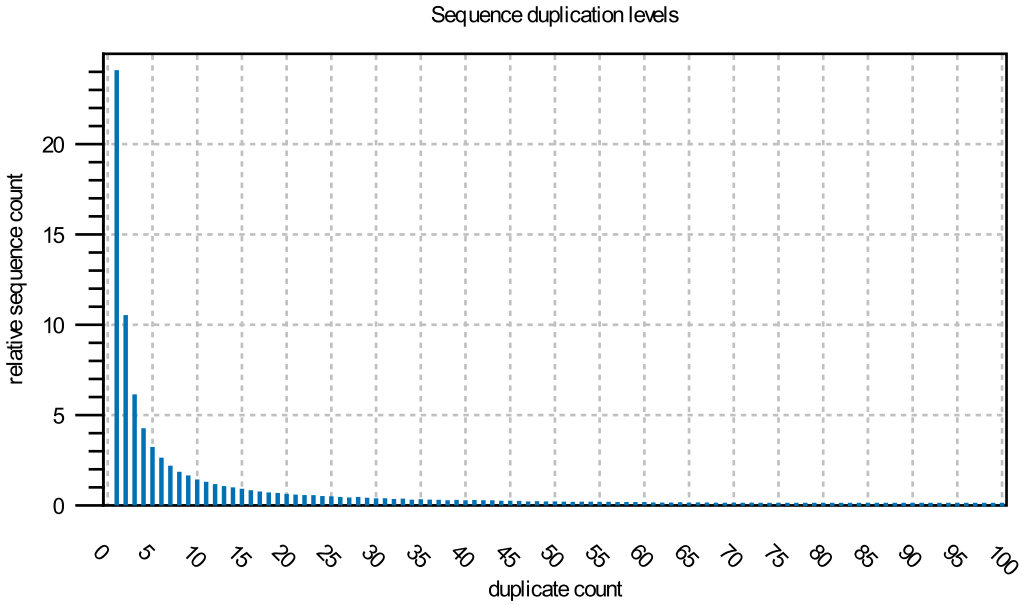
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

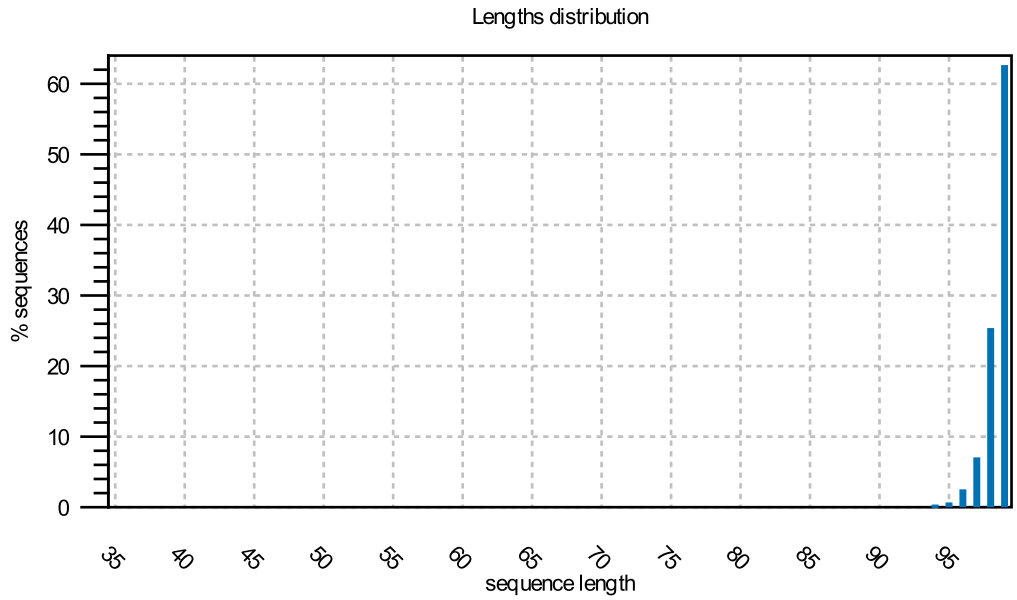
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 13:16:47 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
AAV3_S6:	97,585,420 sequences
Total sequences in data set	97,585,420 sequences
Total nucleotides in data set	9,590,915,338 nucleotides

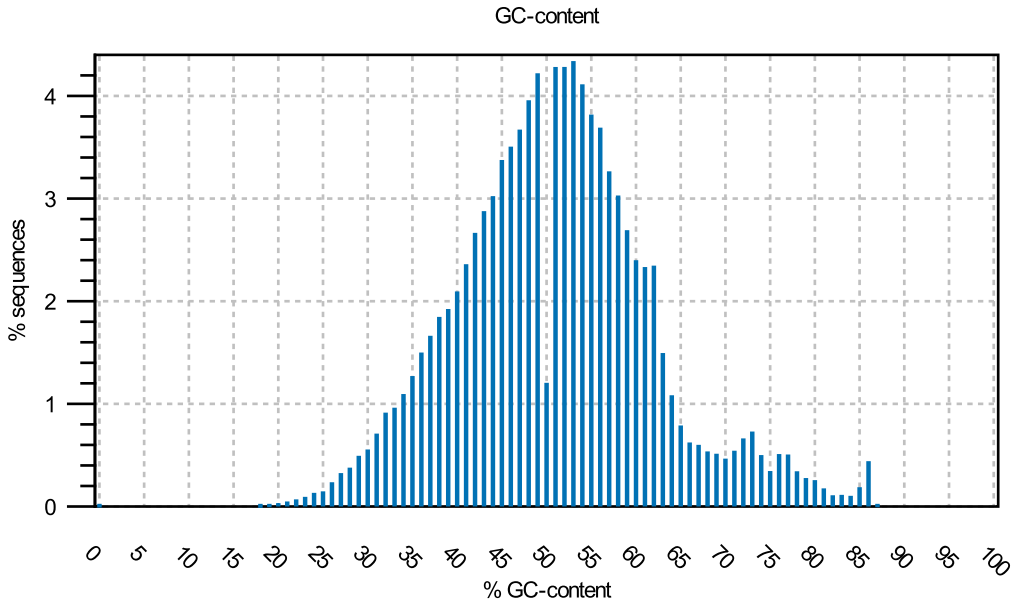
2. Per-sequence analysis

2.1 Lengths distribution



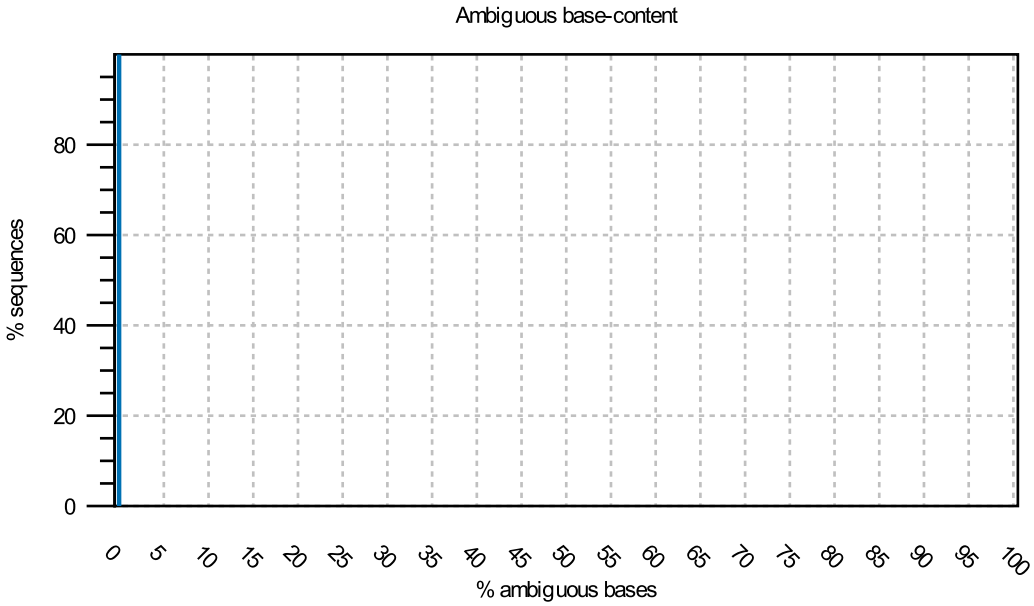
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



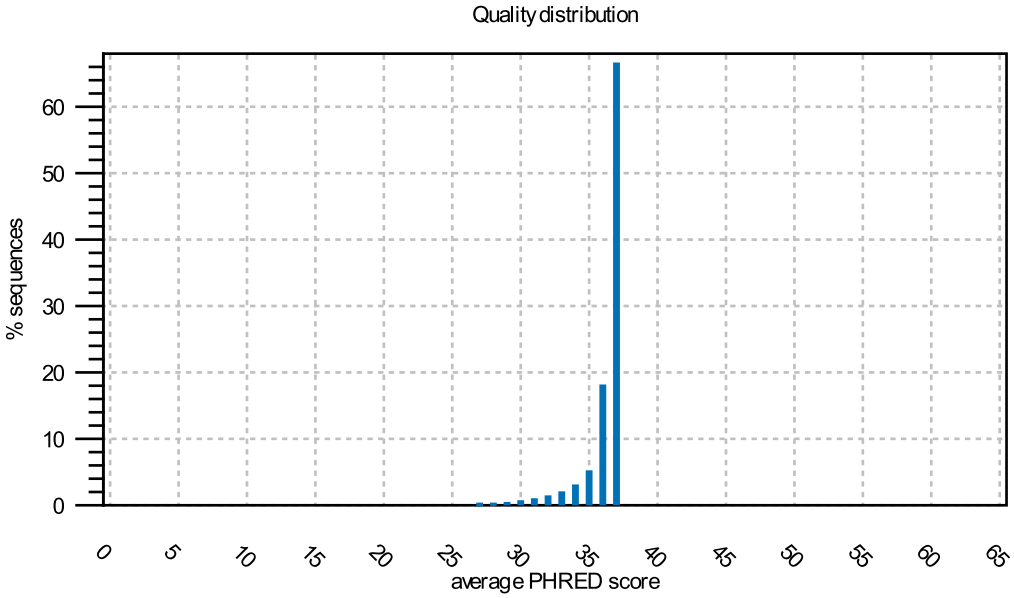
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

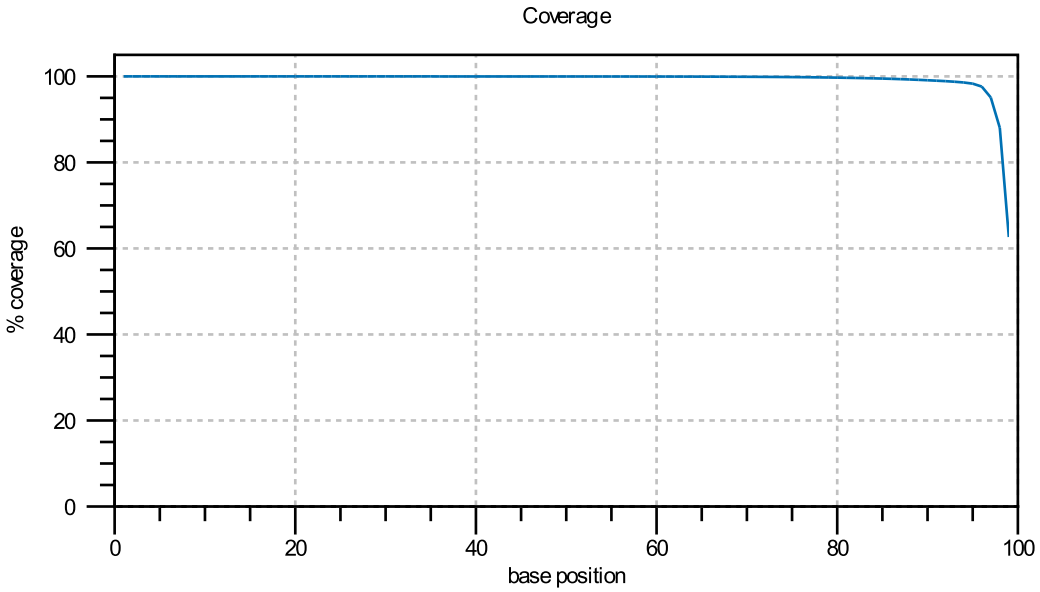
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

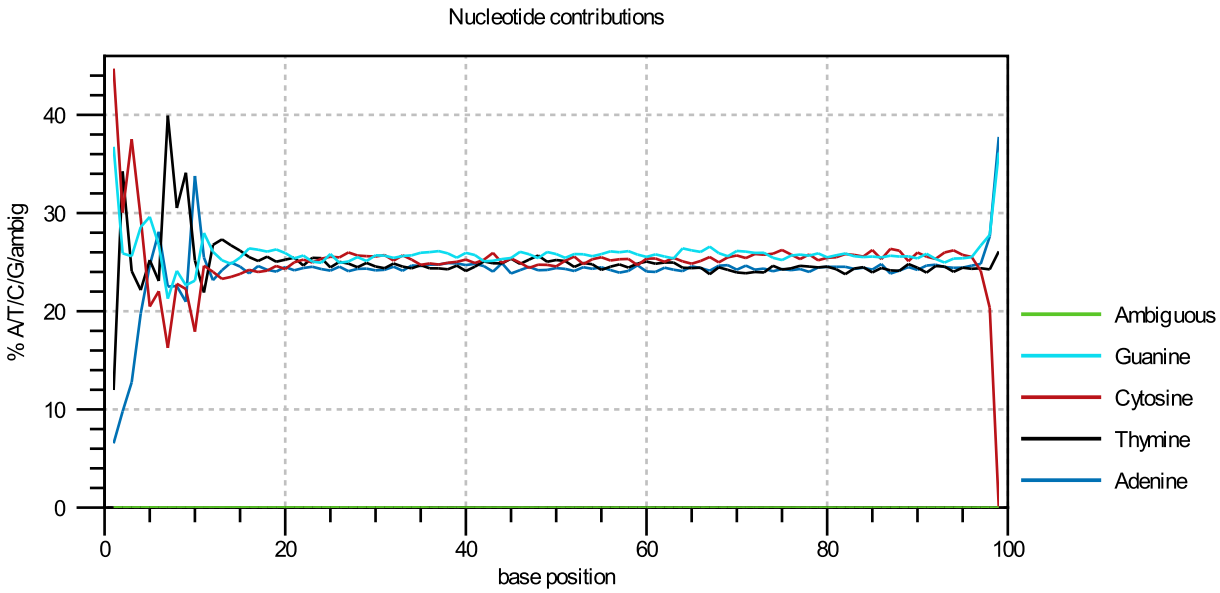
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

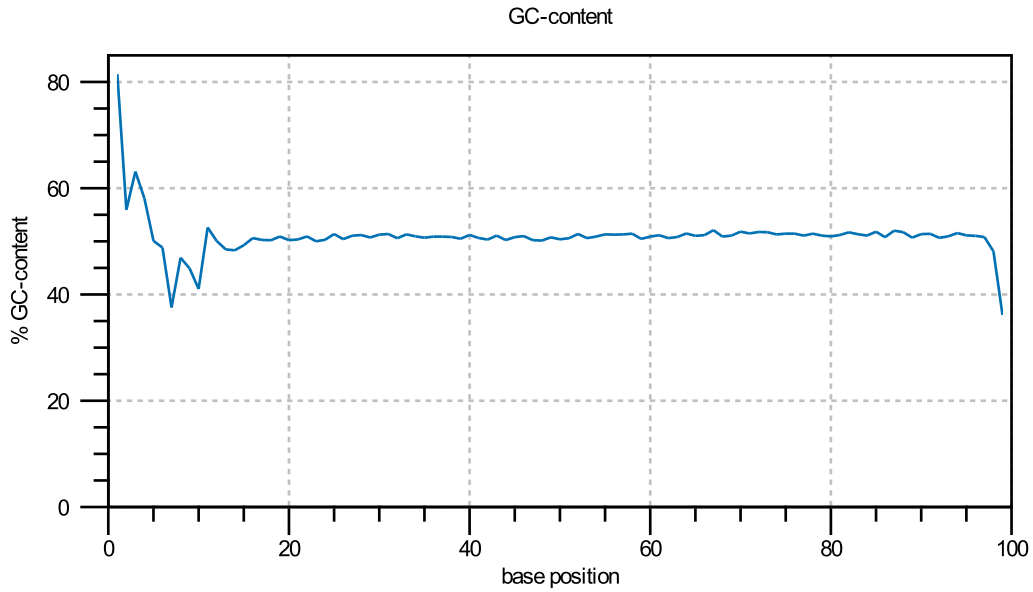
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

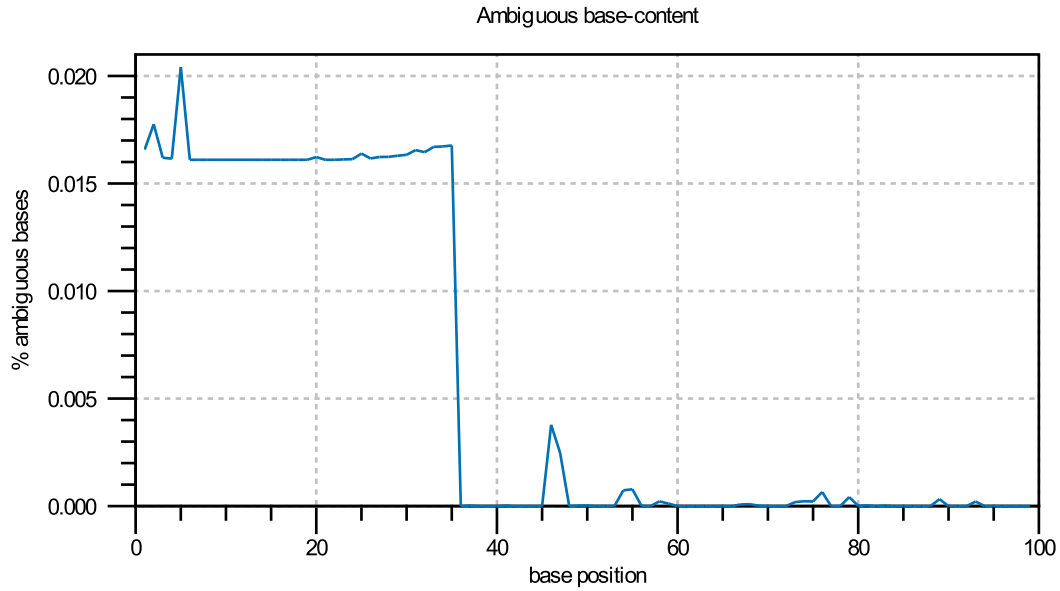
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



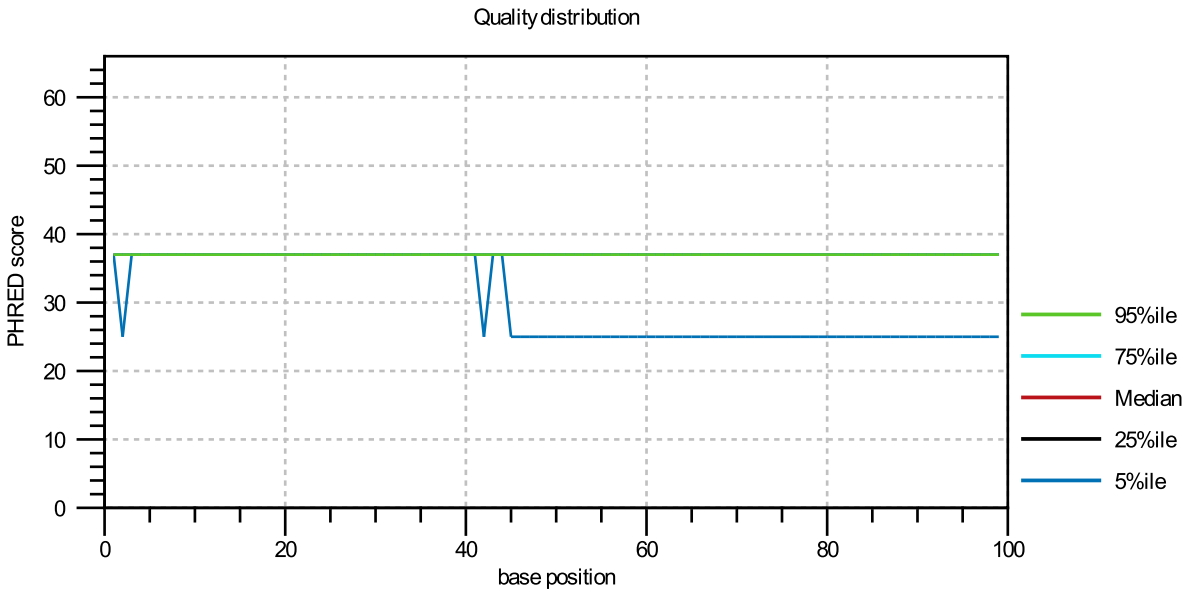
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

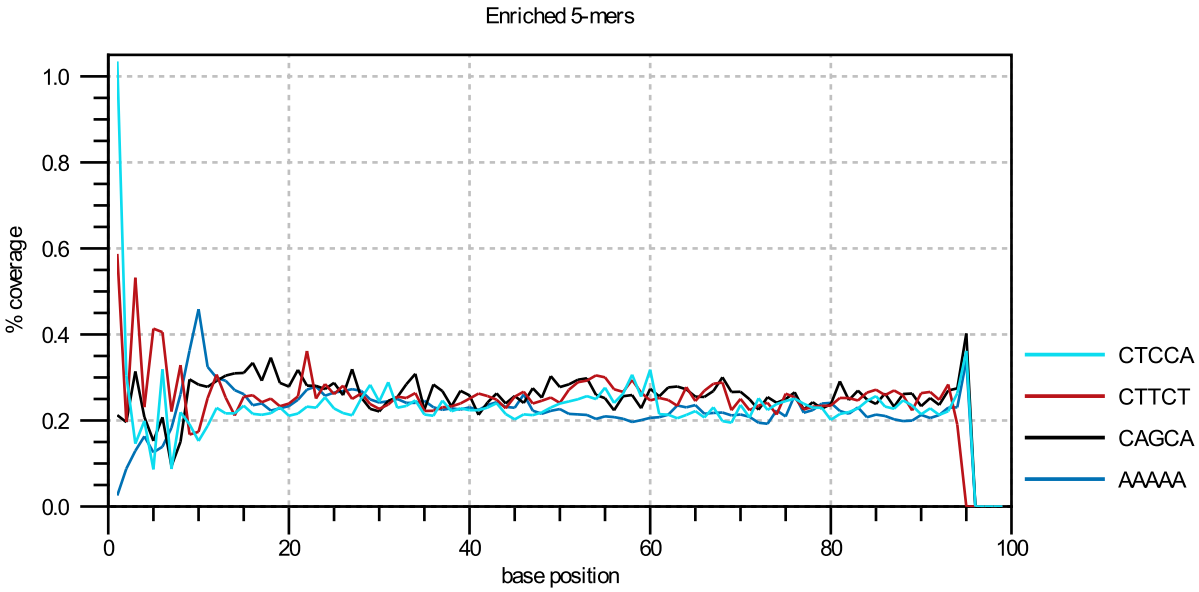
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

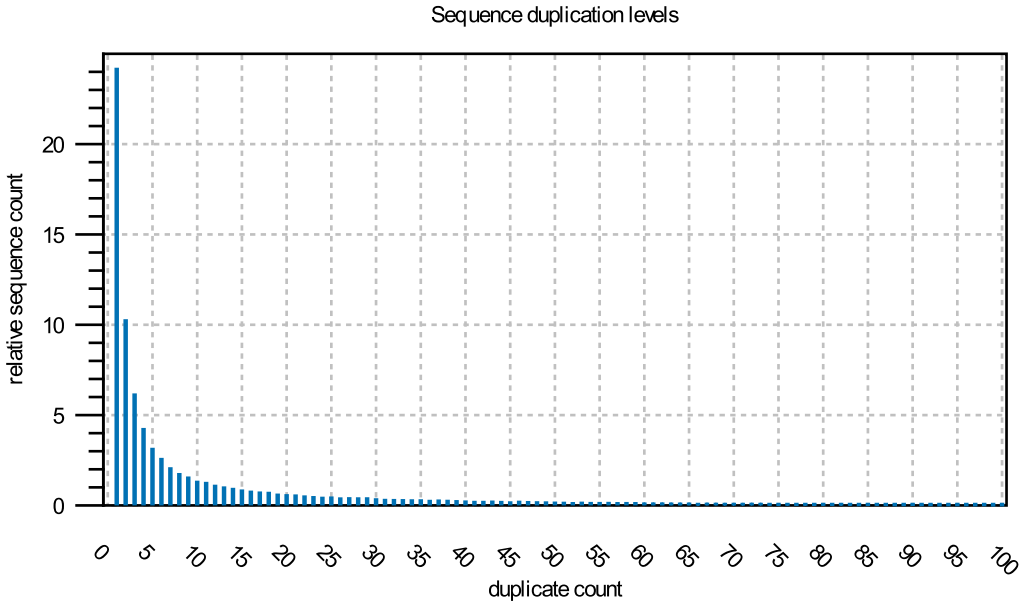
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

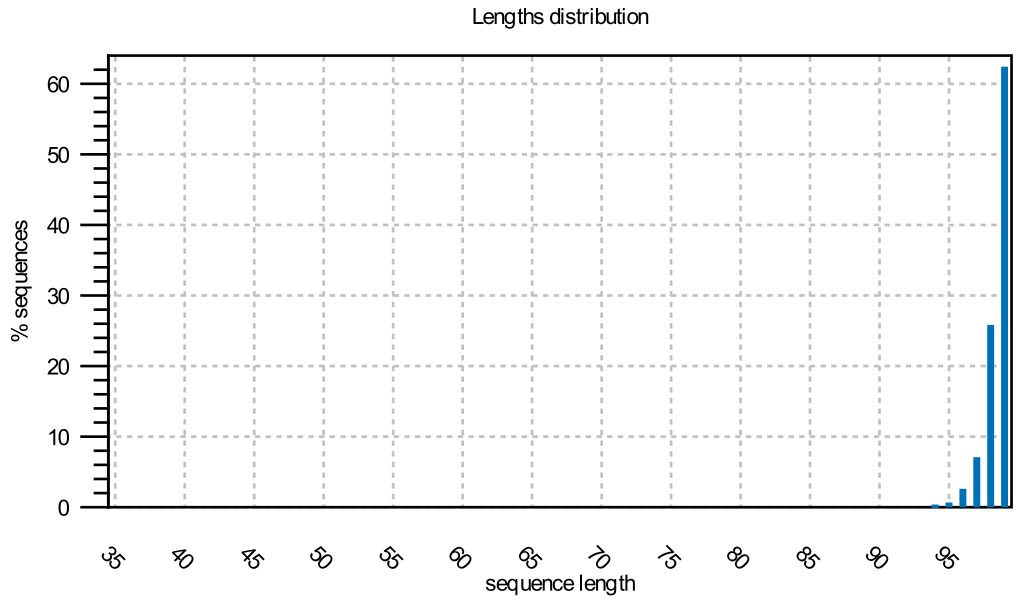
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 15:38:26 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
PB1_S5:	68,966,610 sequences
Total sequences in data set	68,966,610 sequences
Total nucleotides in data set	6,780,920,530 nucleotides

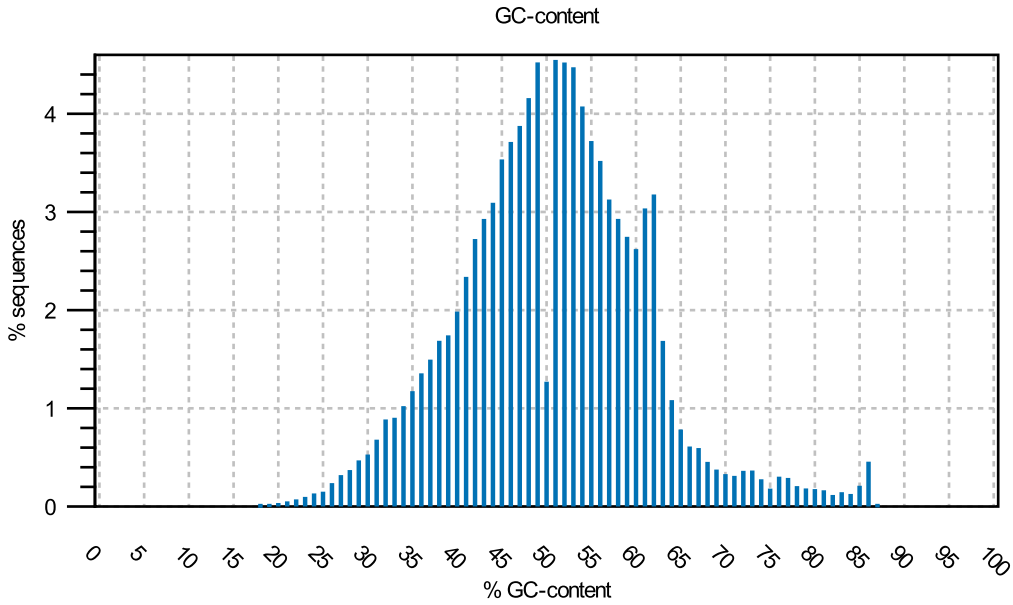
2. Per-sequence analysis

2.1 Lengths distribution



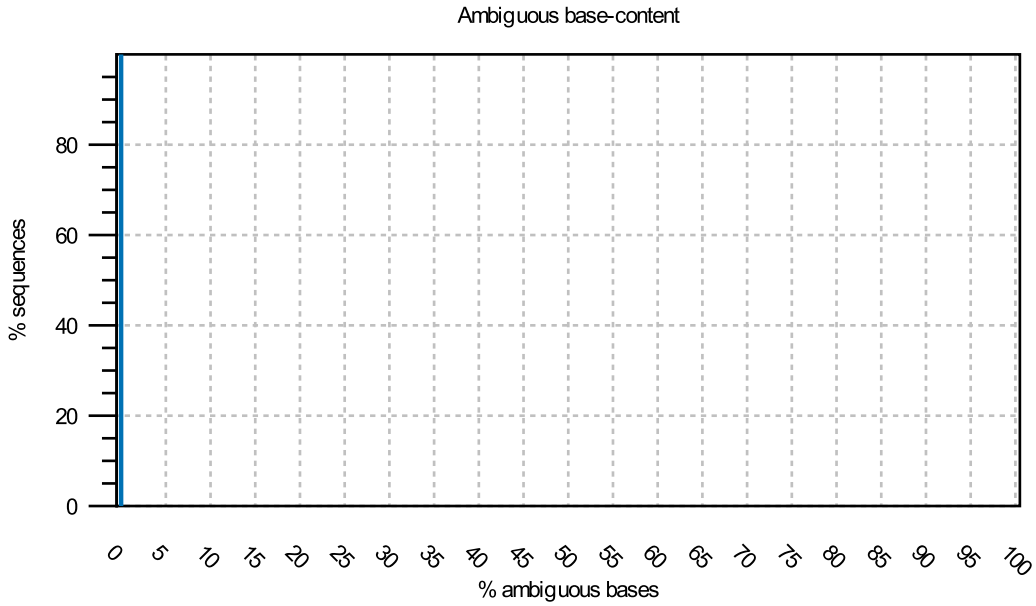
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



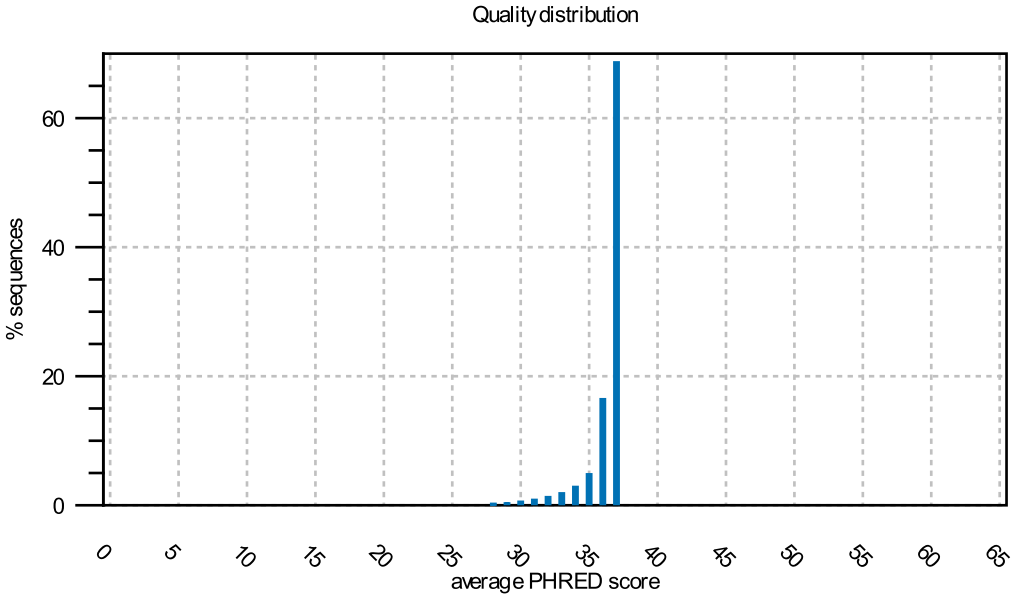
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

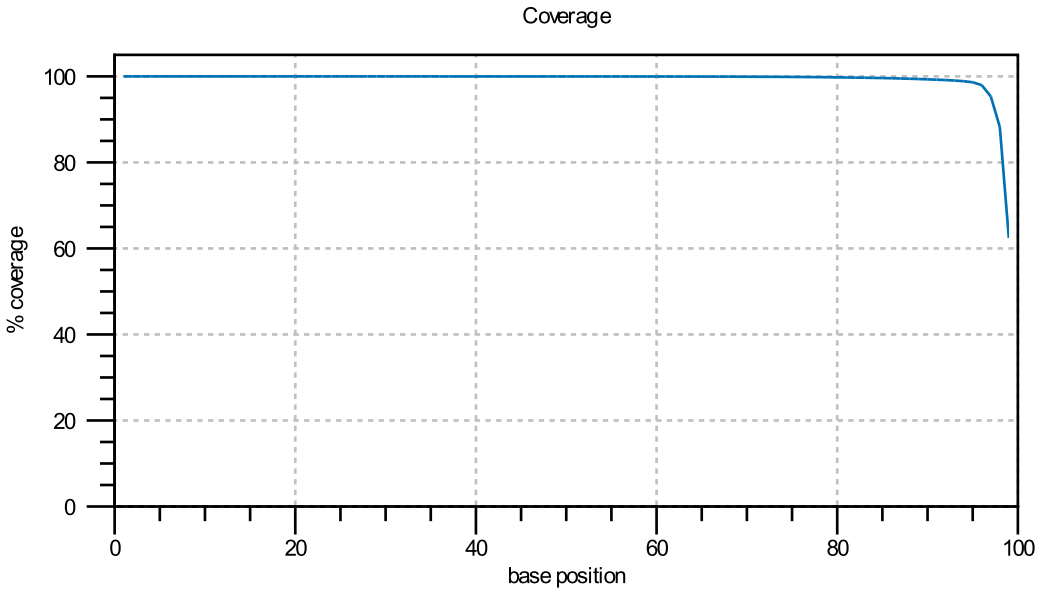
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

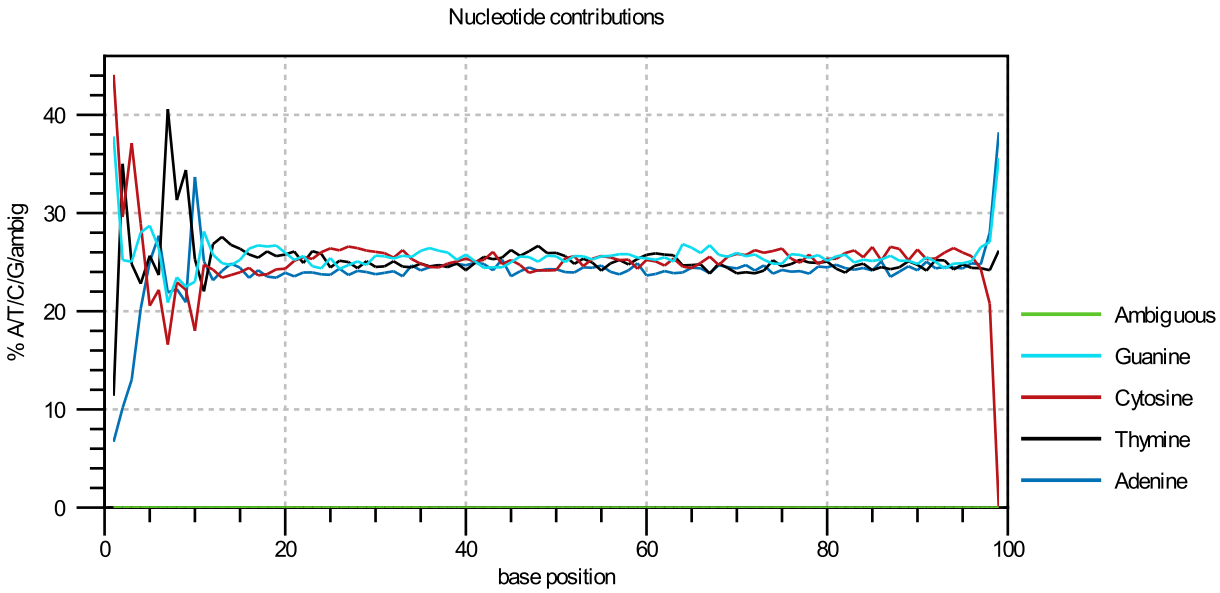
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

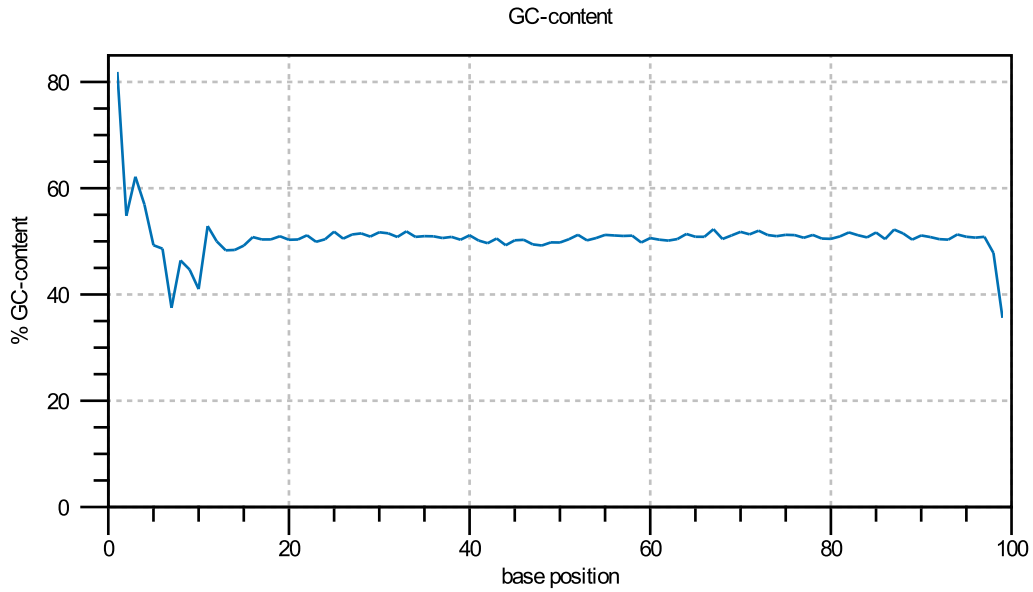
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

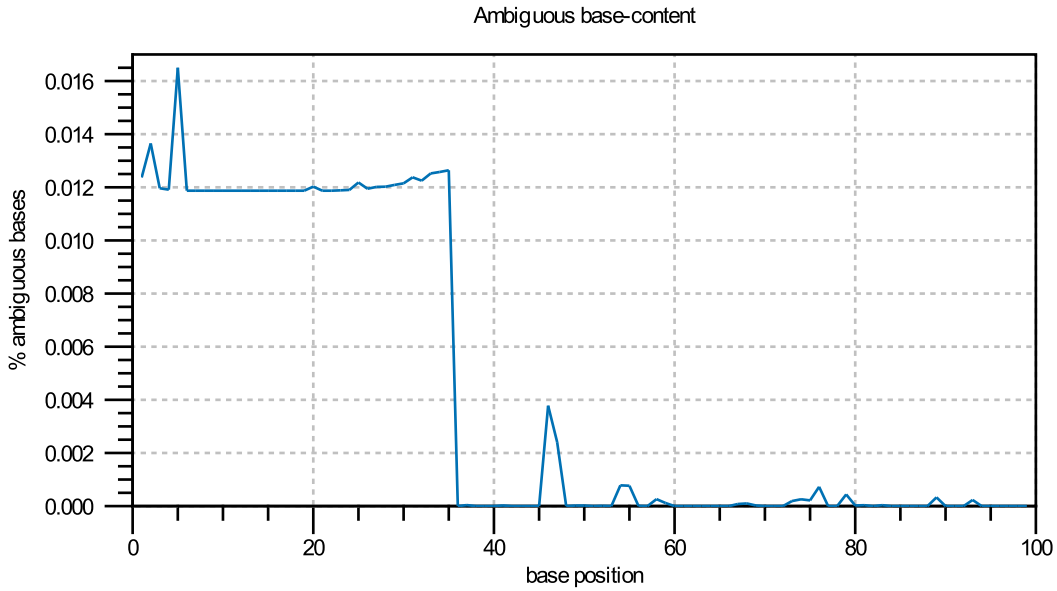
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



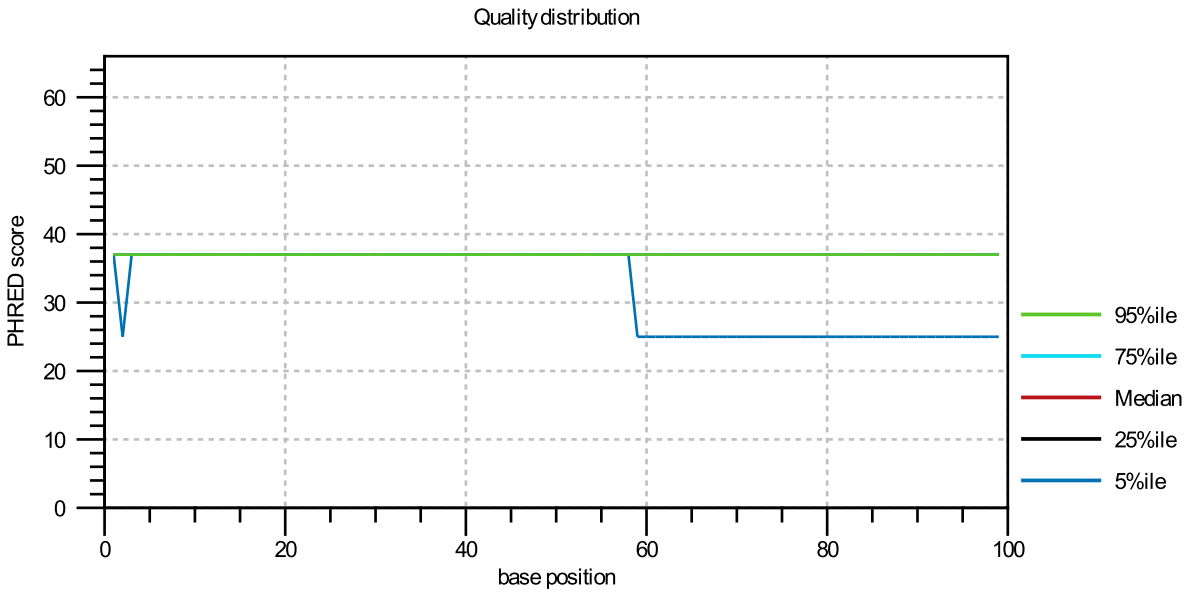
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

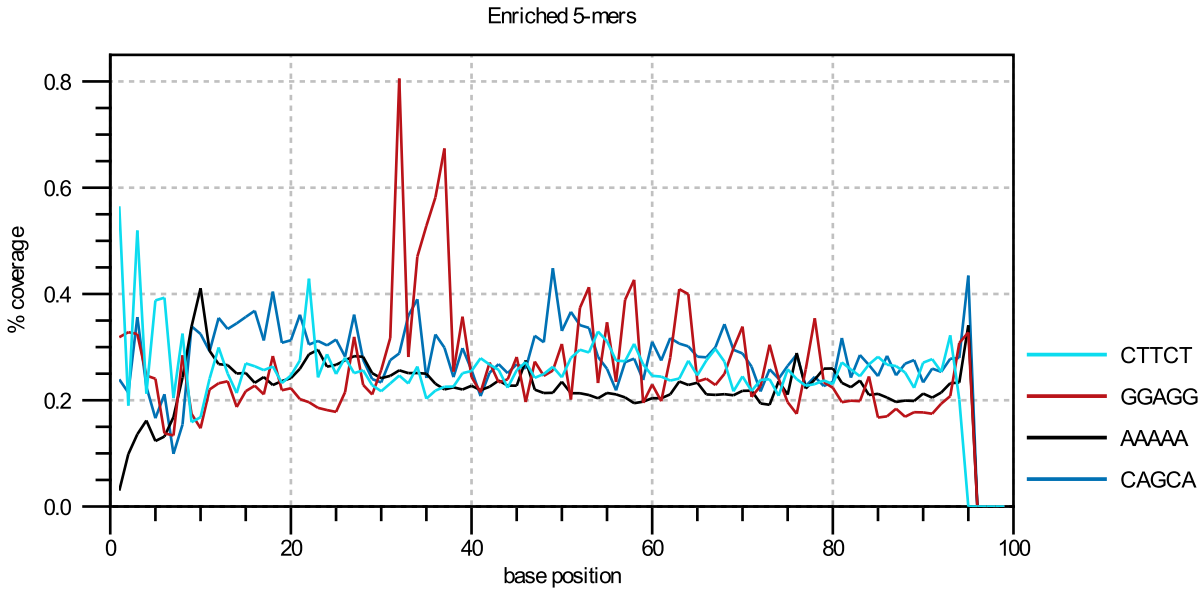
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

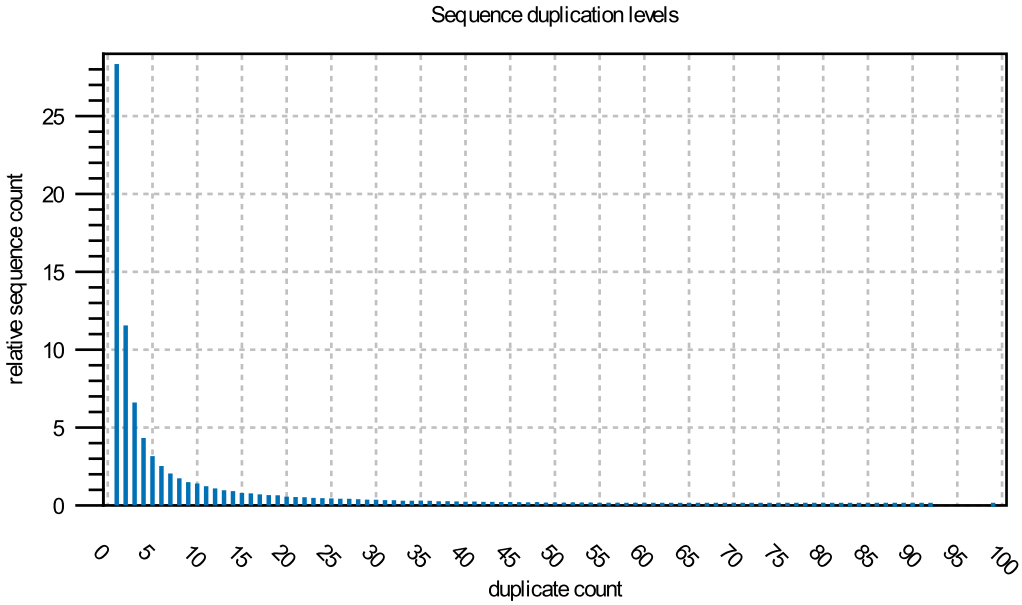
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

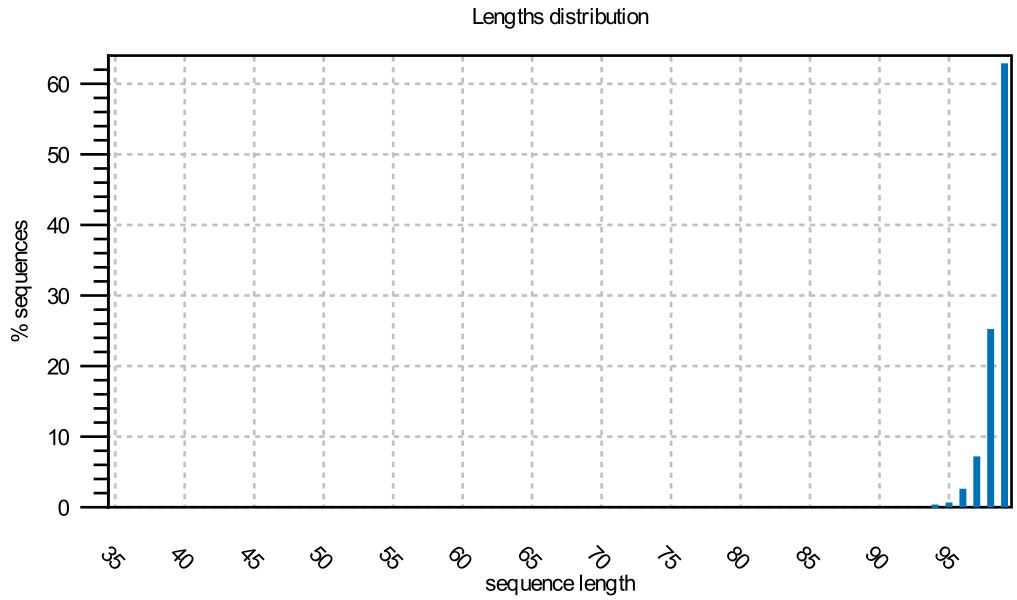
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 15:58:28 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
PB2_S10:	67,920,378 sequences
Total sequences in data set	67,920,378 sequences
Total nucleotides in data set	6,678,602,531 nucleotides

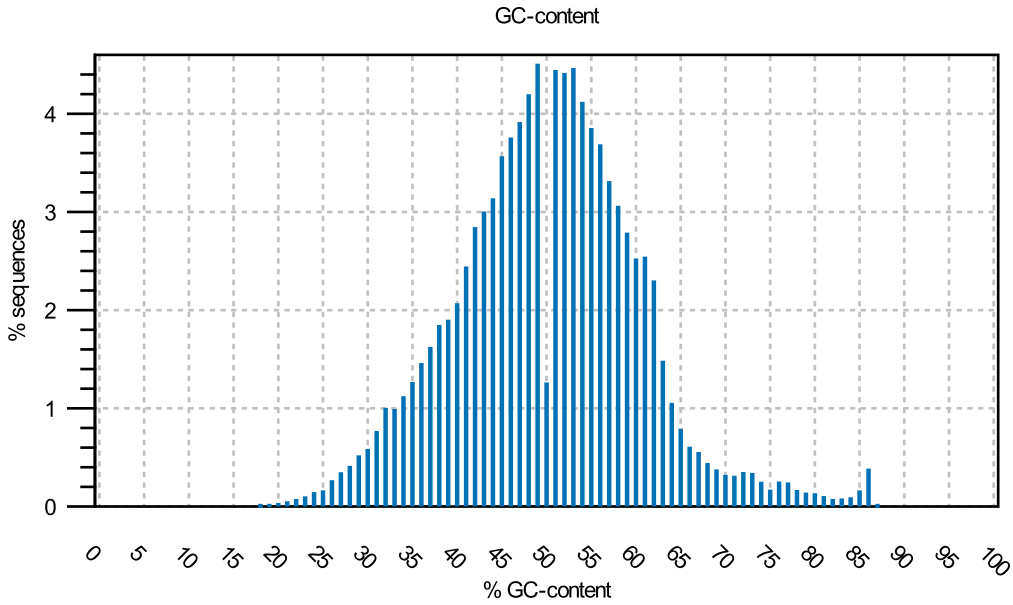
2. Per-sequence analysis

2.1 Lengths distribution



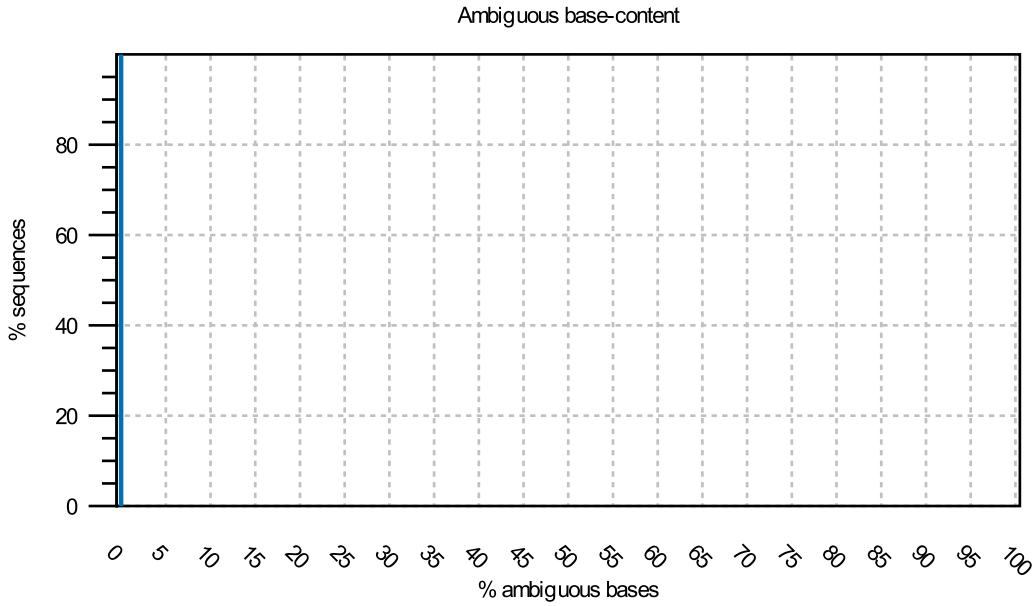
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



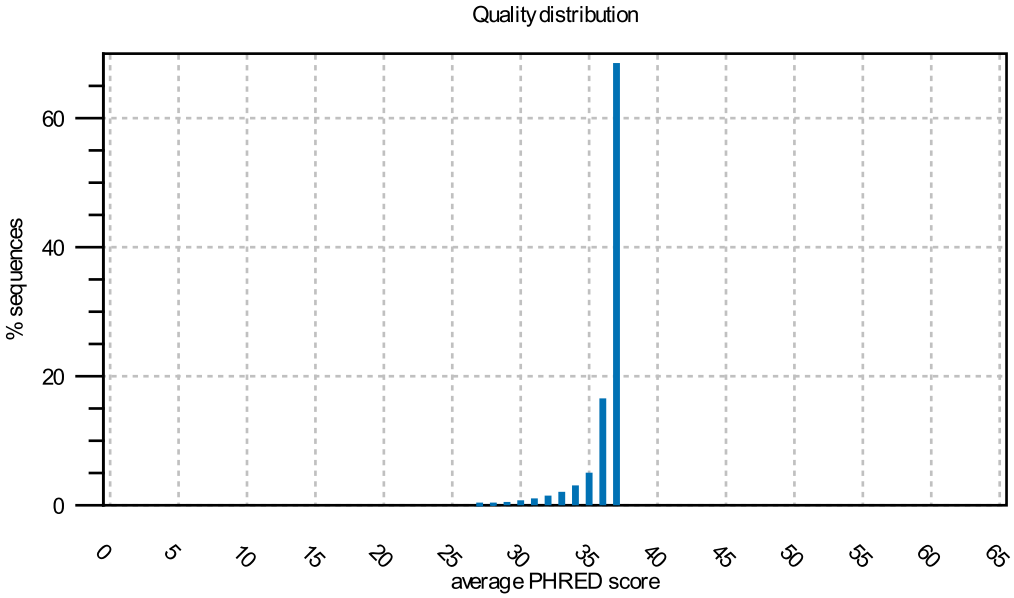
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

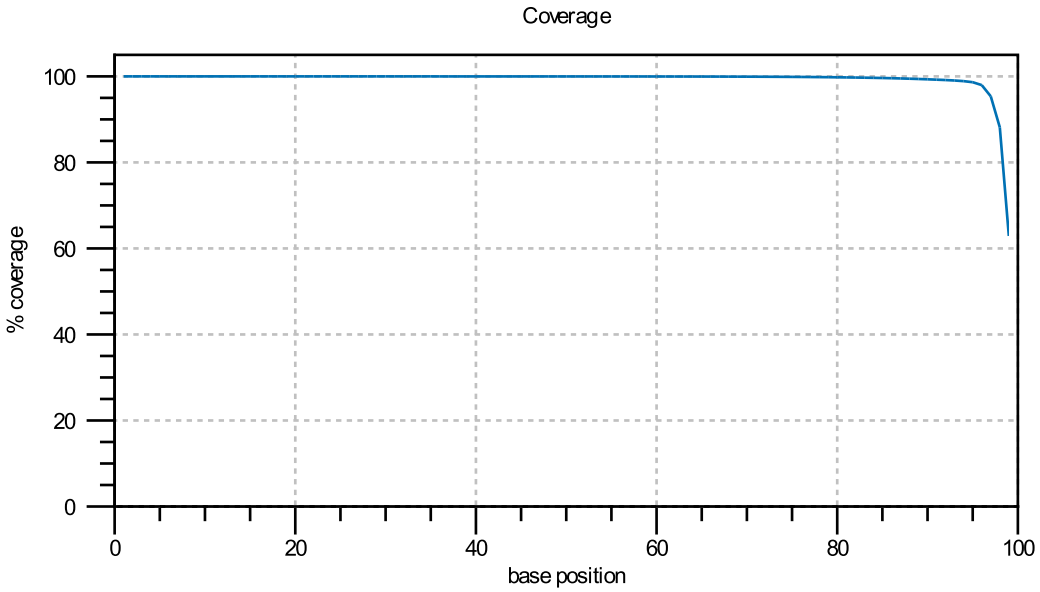
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

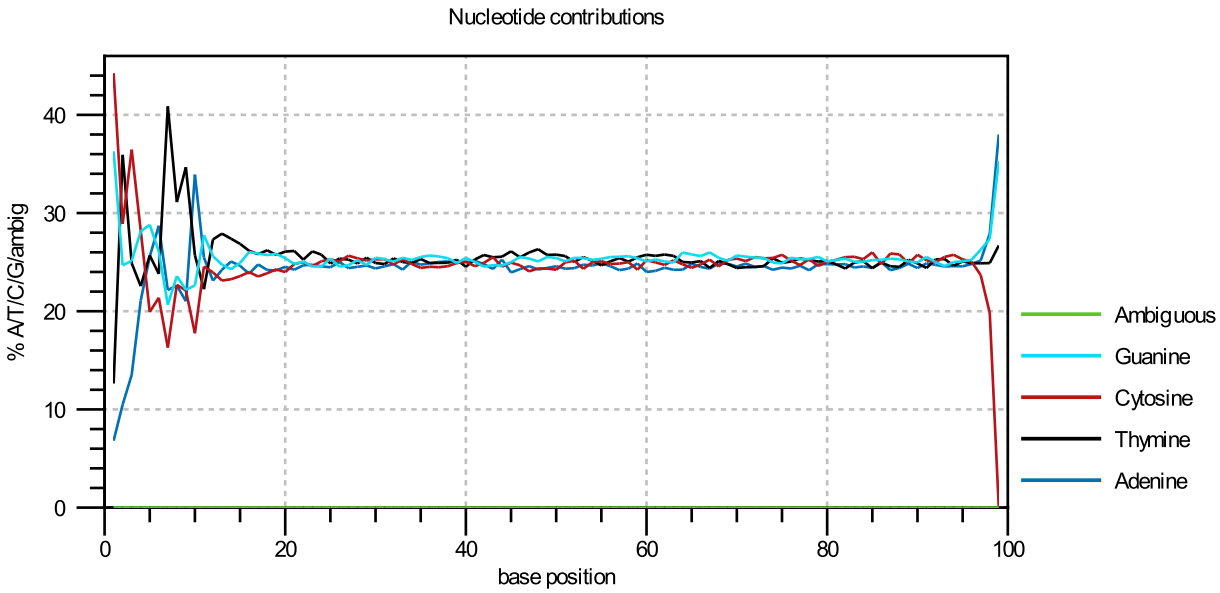
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

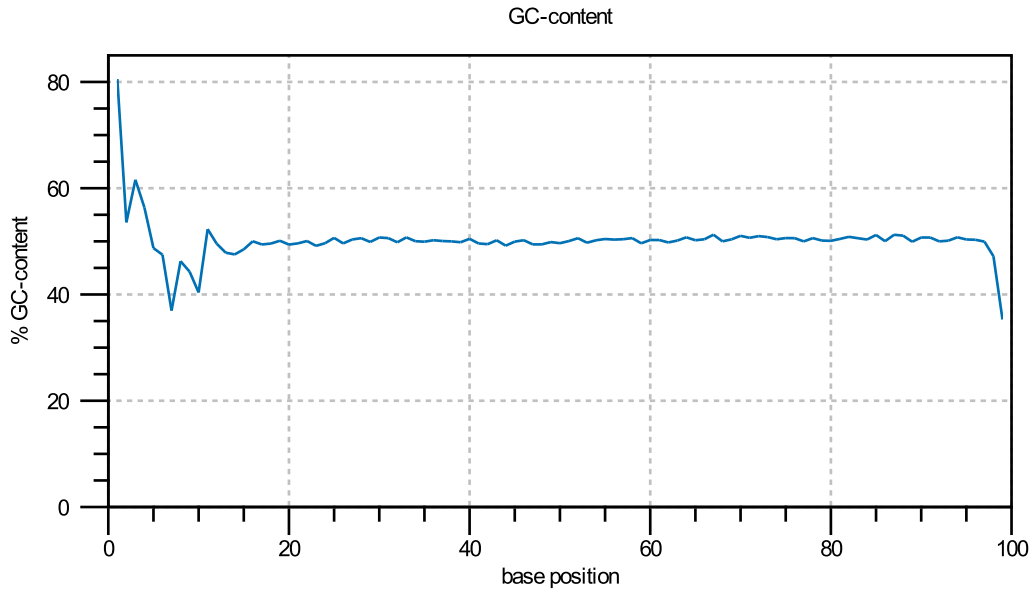
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

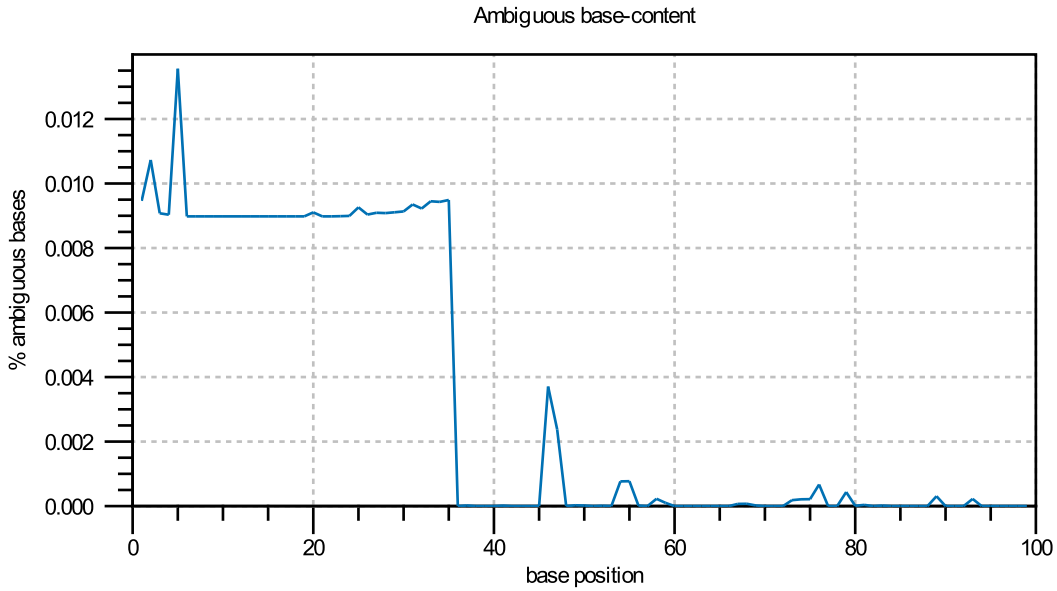
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



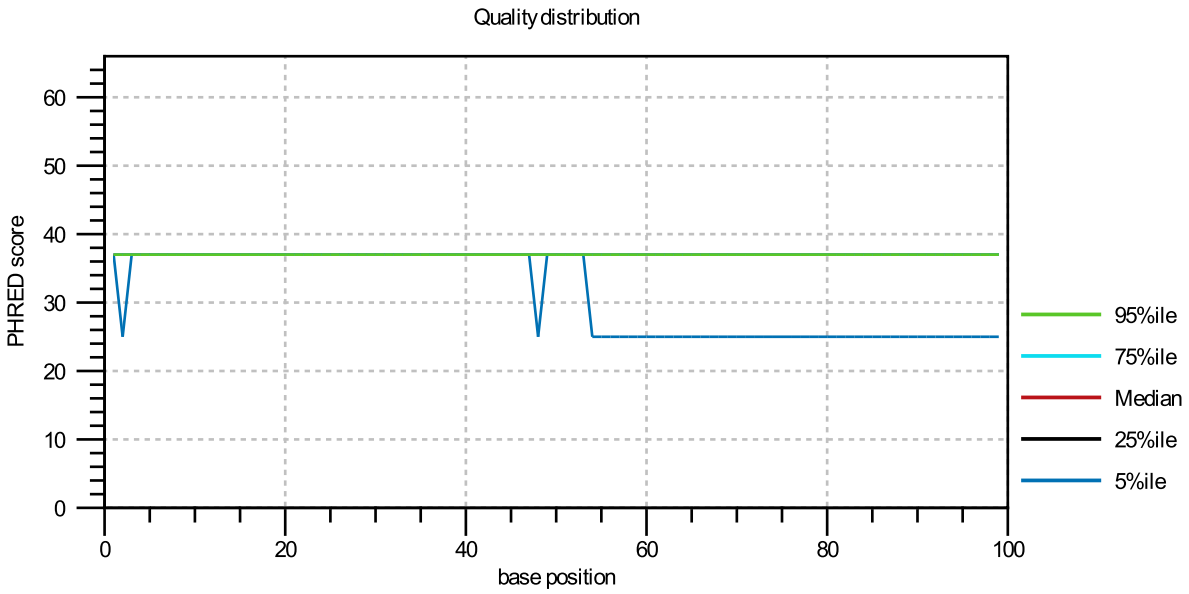
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

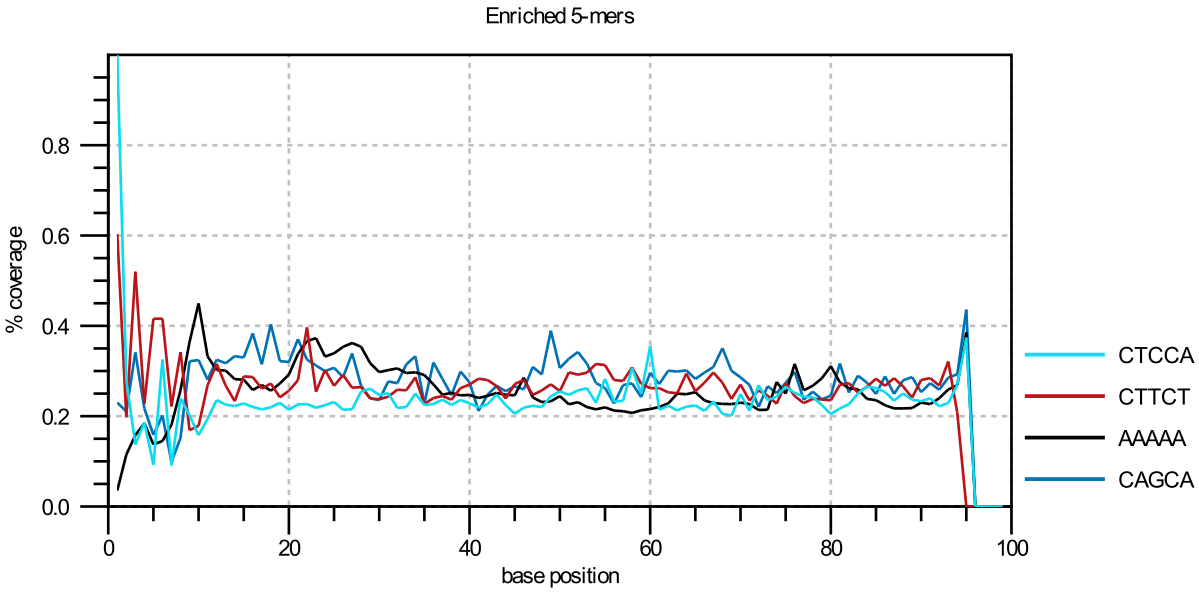
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

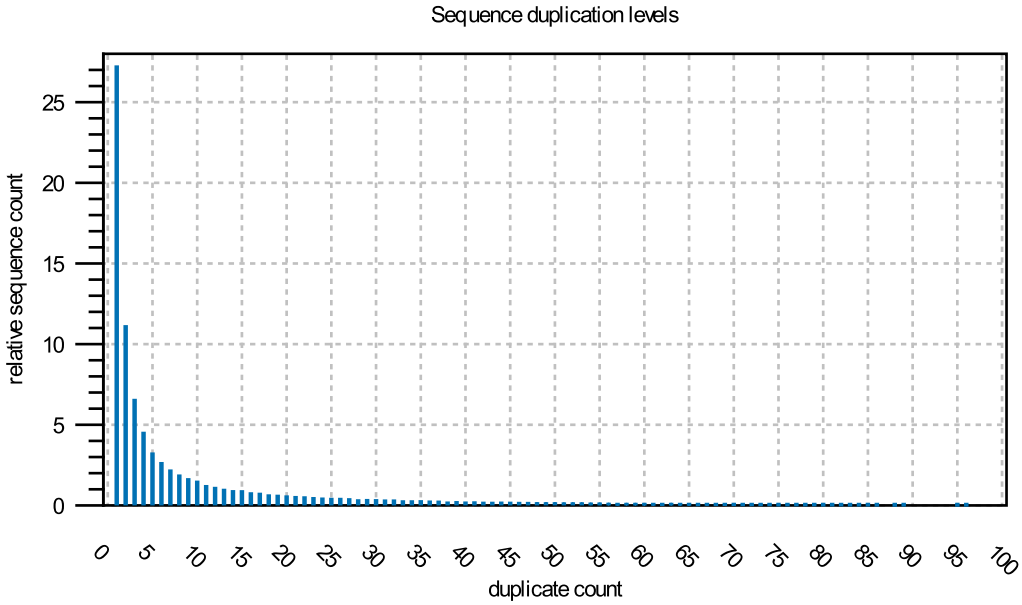
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

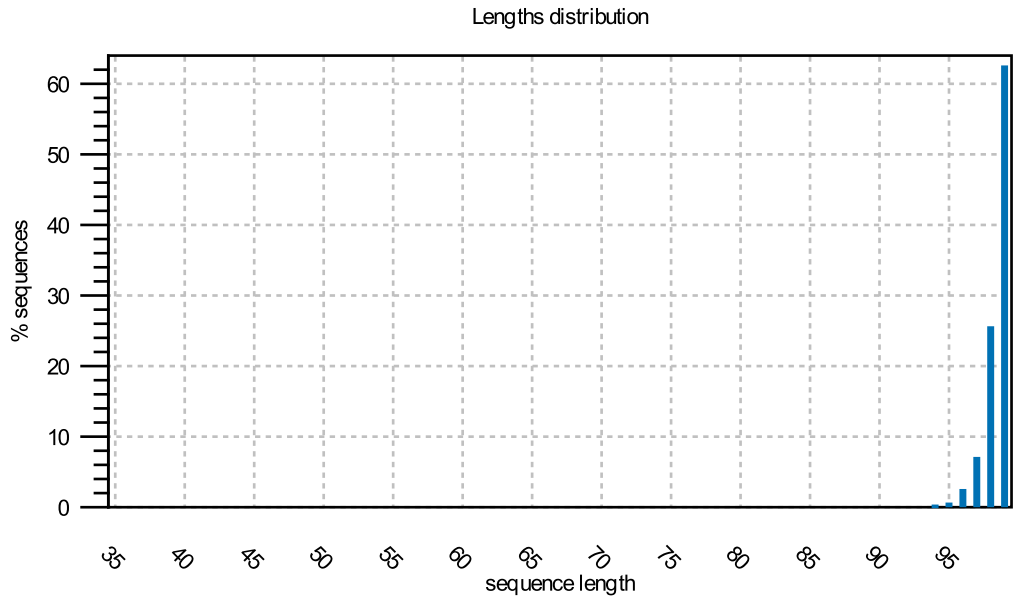
A table of over-represented sequences is given in the supplementary report

1. Summary

Creation date:	Fri Dec 23 16:18:22 AEDT 2022
Generated by:	14048742
Software:	CLC Genomics Workbench 22.0
Based upon:	1 data set
PB3_S8:	80,717,734 sequences
Total sequences in data set	80,717,734 sequences
Total nucleotides in data set	7,936,680,795 nucleotides

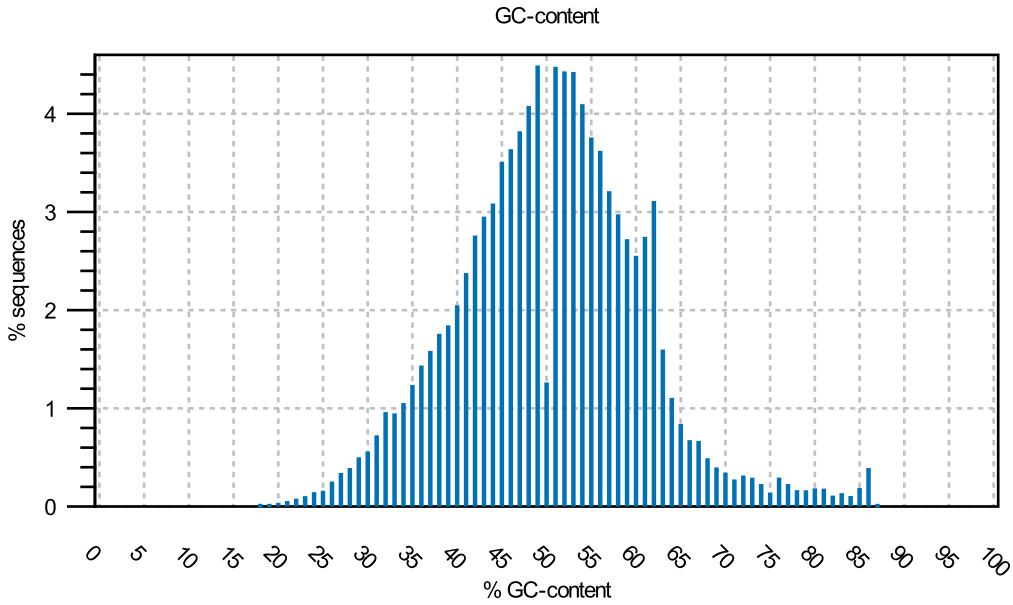
2. Per-sequence analysis

2.1 Lengths distribution



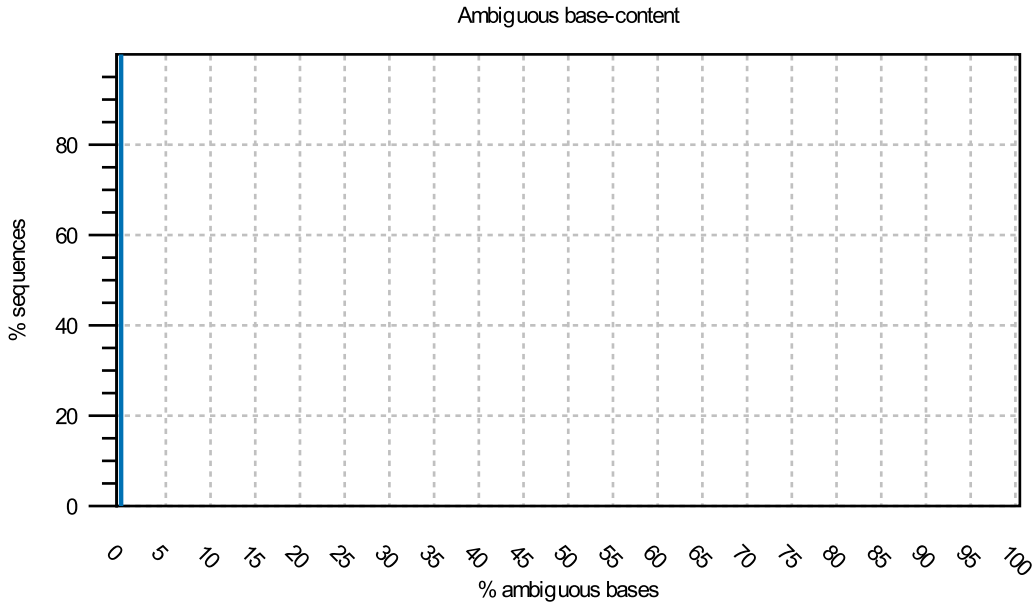
Distribution of sequence lengths. In cases of untrimmed Illumina reads it will just contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of sequences

2.2 GC-content



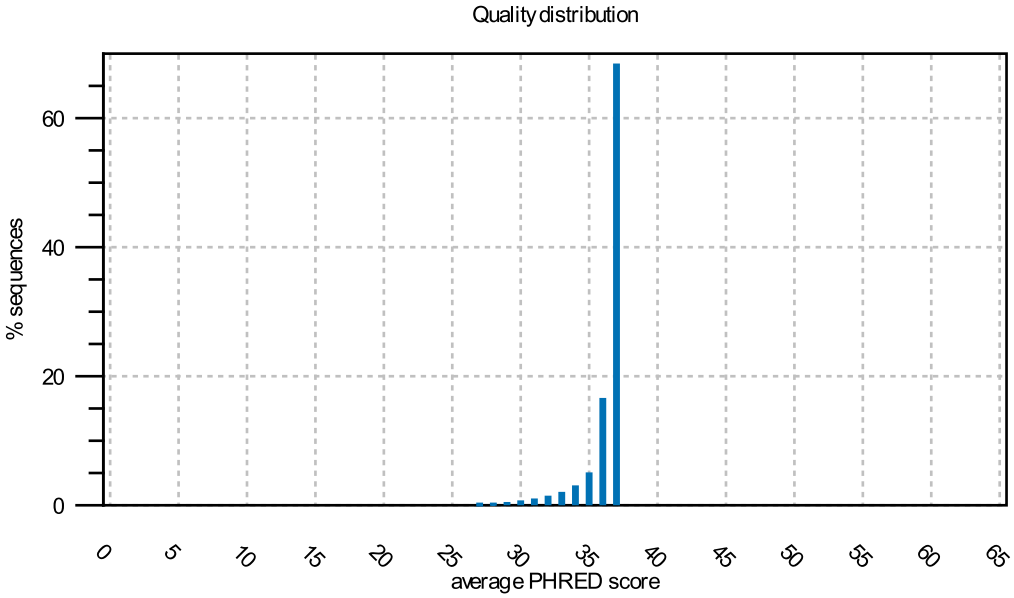
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of GC-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of ambiguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number of sequences

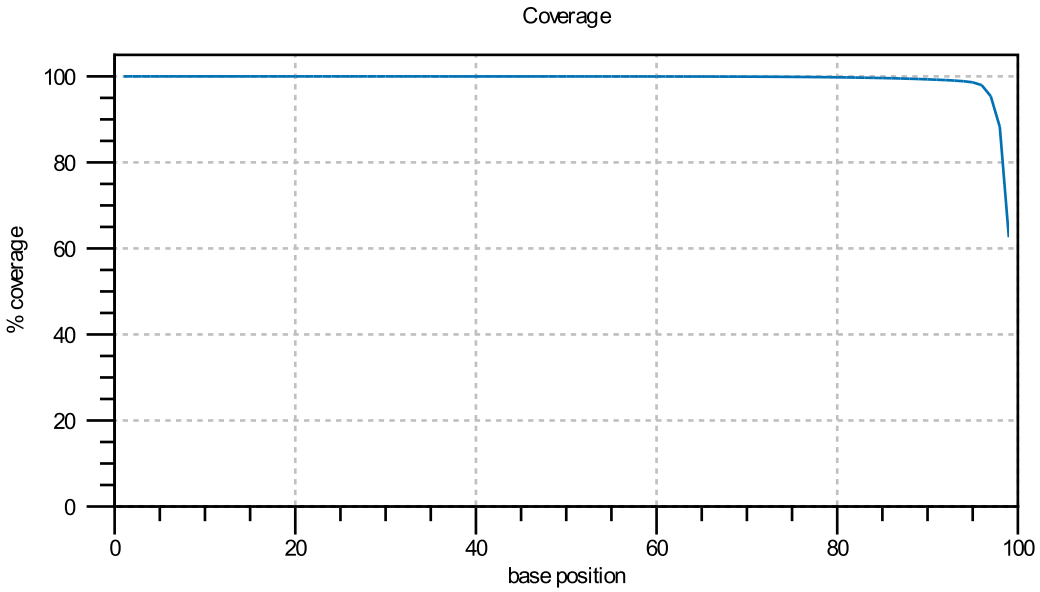
2.4 Quality distribution



Distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequences

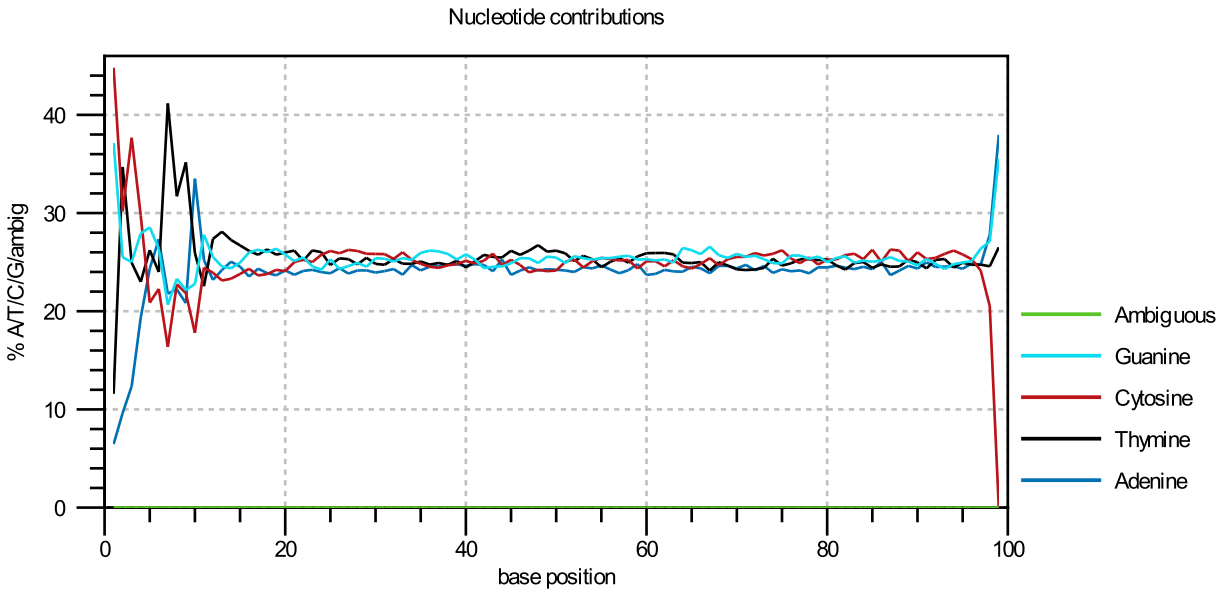
3. Per-base analysis

3.1 Coverage



The number of sequences that support (cover) the individual base positions. In cases of untrimmed Illumina reads it will just contain a line.
x: base position
y: number of sequences covering individual base positions normalized to the total number of sequences

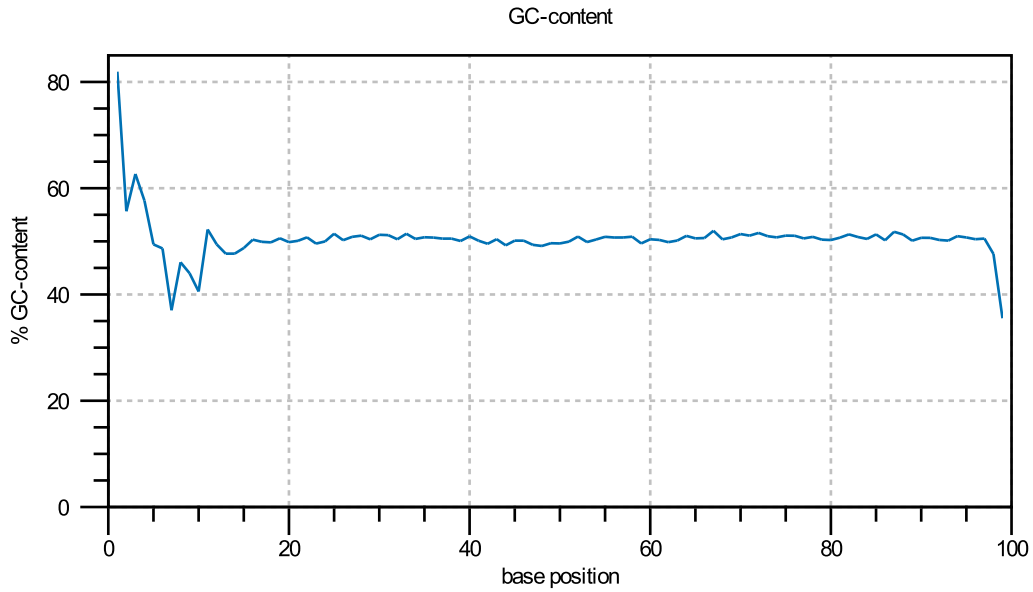
3.2 Nucleotide contributions



Coverages for the four DNA nucleotides and ambiguous bases.
x: base position

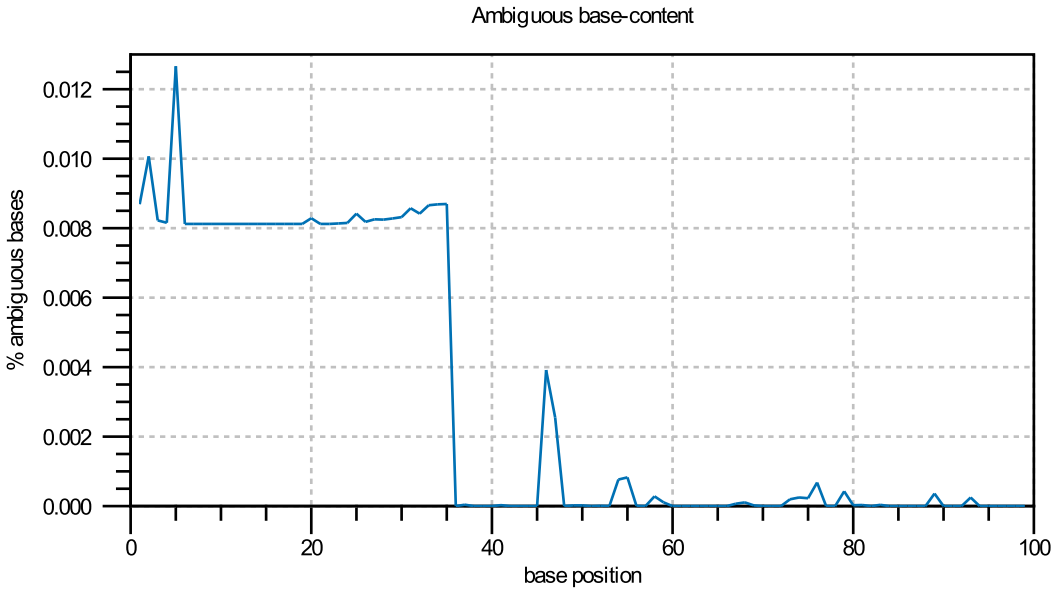
y: number of nucleotides observed per type normalized to the total number of nucleotides observed at that position

3.3 GC-content



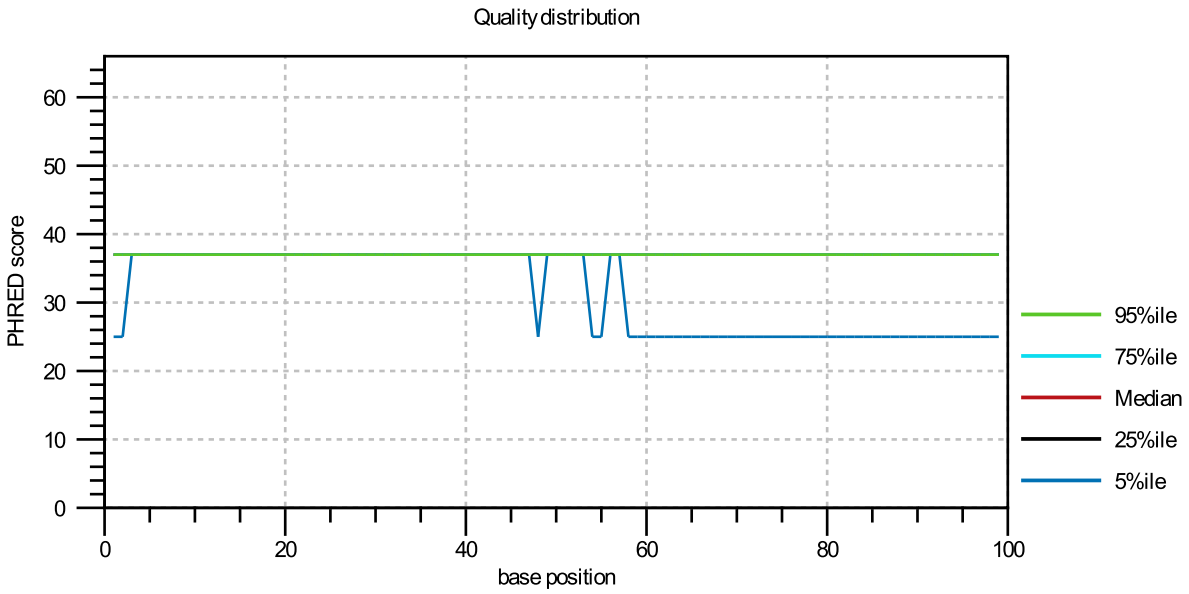
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number of bases observed at that position

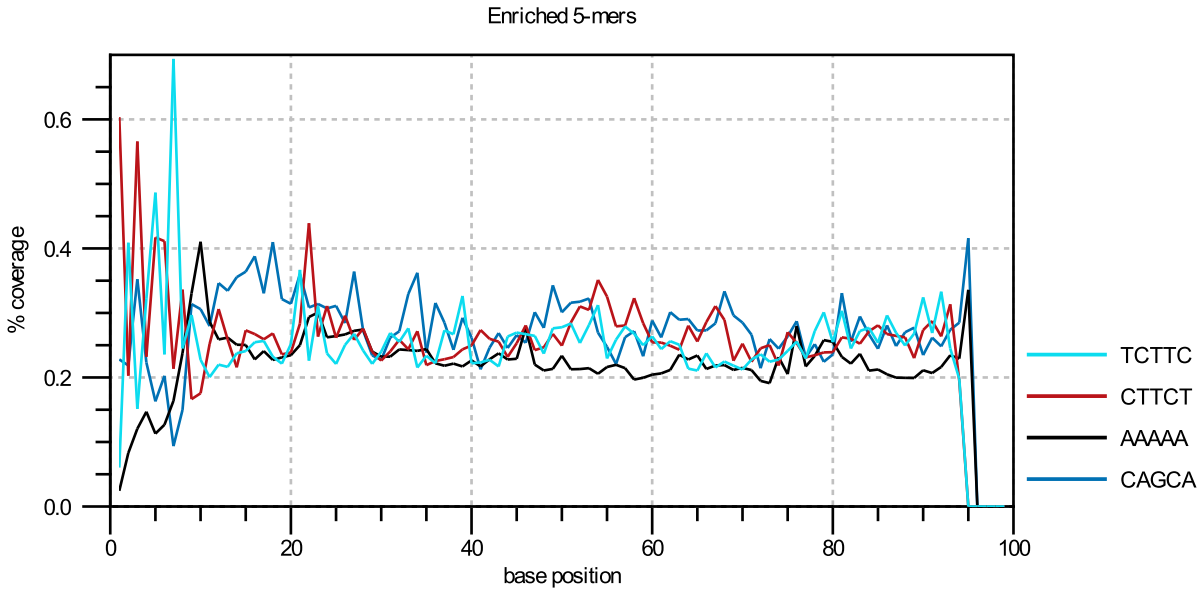
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

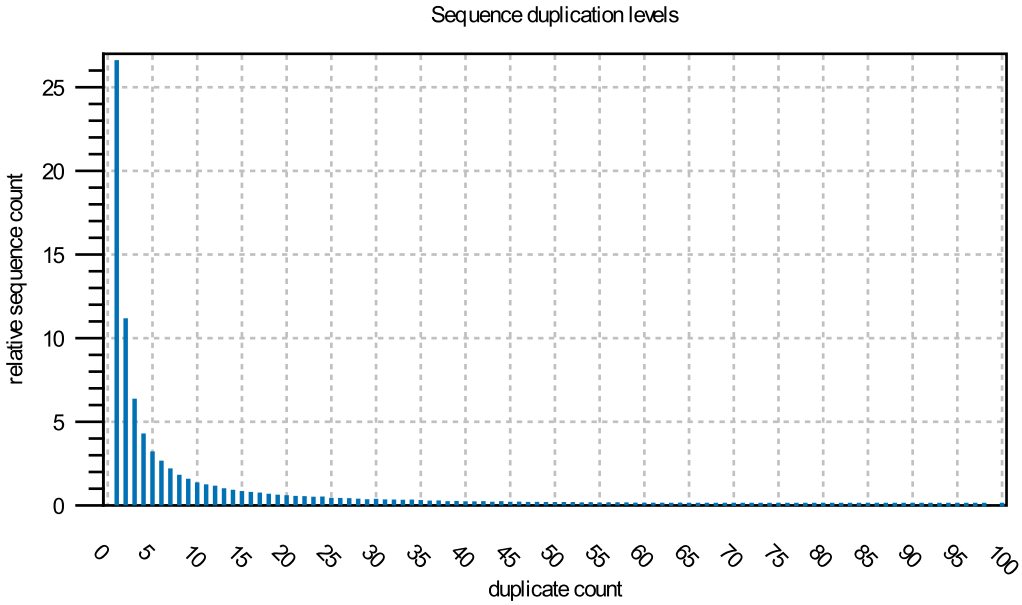
4. Over-representation analyses

4.1 Enriched 5-mers



The five most-overrepresented 5-mers. The over-representation of a 5-mer is calculated as the ratio of the observed and expected 5-mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5-mer. (5-mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5-mer has been observed normalized to all 5-mers observed at that position

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a particular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of unique sequences

4.3 Duplicated sequences

A table of over-represented sequences is given in the supplementary report

Table of contents

1. RNA-Seq analysis 3
    1.1 Read count statistics 3
    1.2 Fragment counting statistics 4
    1.3 Spike-in quality control 5
    1.4 Strand specificity 5
    1.5 Adapter read-through 6
    1.6 Distribution of biotypes 6
    1.7 Transcript length coverage 13

1. RNA-Seq analysis

1.1 Read count statistics

For paired data, there are two reads in a pair.

The table is based on 12 samples.

Sample name	Read count	Single, mapped (%)	Single, not mapped (%)	Paired, mapped pairs (%)
NL1_S1 (RNA-Seq report)	71,177,957	98.19	1.81	-
NL2_S11 (RNA-Seq report)	78,295,601	98.57	1.43	-
NL3_S4 (RNA-Seq report)	75,867,324	97.82	2.18	-
DL1_S7 (RNA-Seq report)	71,673,917	97.80	2.20	-
DL2_S3 (RNA-Seq report)	78,503,589	98.12	1.88	-
DL3_S2 (RNA-Seq report)	75,928,949	97.89	2.11	-
AAV1_S12 (RNA-Seq report)	73,054,598	97.71	2.29	-
AAV2_S9 (RNA-Seq report)	88,511,361	97.53	2.47	-
AAV3_S6 (RNA-Seq report)	97,585,420	96.67	3.33	-
PB1_S5 (RNA-Seq report)	68,966,610	97.82	2.18	-
PB2_S10 (RNA-Seq report)	67,920,378	98.13	1.87	-
PB3_S8 (RNA-Seq report)	80,717,734	98.08	1.92	-
Minimum	67,920,378.00	96.67	1.43	-
Median	75,898,136.50	97.85	2.15	-
Maximum	97,585,420.00	98.57	3.33	-
Mean	77,350,286.50	97.86	2.14	-
Standard deviation	8,526,930.53	0.46	0.46	-

Sample name	Paired, broken pairs (%)	Paired, not mapped (%)
NL1_S1 (RNA-Seq report)	-	-

Sample name	Paired, broken pairs (%)	Paired, not mapped (%)
NL2_S11 (RNA-Seq report)	-	-
NL3_S4 (RNA-Seq report)	-	-
DL1_S7 (RNA-Seq report)	-	-
DL2_S3 (RNA-Seq report)	-	-
DL3_S2 (RNA-Seq report)	-	-
AAV1_S12 (RNA-Seq report)	-	-
AAV2_S9 (RNA-Seq report)	-	-
AAV3_S6 (RNA-Seq report)	-	-
PB1_S5 (RNA-Seq report)	-	-
PB2_S10 (RNA-Seq report)	-	-
PB3_S8 (RNA-Seq report)	-	-
Minimum	-	-
Median	-	-
Maximum	-	-
Mean	-	-
Standard deviation	-	-

1.2 Fragment counting statistics

The table is based on 12 samples.

Sample name	Mapped to genes (%)	Mapped to intergenic (%)
NL1_S1 (RNA-Seq report)	96.58	3.42
NL2_S11 (RNA-Seq report)	96.69	3.31
NL3_S4 (RNA-Seq report)	95.99	4.01
DL1_S7 (RNA-Seq report)	96.25	3.75
DL2_S3 (RNA-Seq report)	95.94	4.06
DL3_S2 (RNA-Seq report)	96.42	3.58
AAV1_S12 (RNA-Seq report)	96.25	3.75
AAV2_S9 (RNA-Seq report)	95.84	4.16
AAV3_S6 (RNA-Seq report)	95.93	4.07

Sample name	Mapped to genes (%)	Mapped to intergenic (%)
PB1_S5 (RNA-Seq report)	96.24	3.76
PB2_S10 (RNA-Seq report)	96.59	3.41
PB3_S8 (RNA-Seq report)	96.14	3.86
Minimum	95.84	3.31
Median	96.24	3.76
Maximum	96.69	4.16
Mean	96.24	3.76
Standard deviation	0.28	0.28

1.3 Spike-in quality control

No data available

1.4 Strand specificity

The table is based on 12 samples.

Sample name	Strand specific setting	Forward reads mapped (%)	Reverse reads mapped (%)	Ignored reads (wrong strand) (%)
NL1_S1 (RNA-Seq report)	Both	1.59	98.41	0
NL2_S11 (RNA-Seq report)	Both	1.17	98.83	0
NL3_S4 (RNA-Seq report)	Both	1.71	98.29	0
DL1_S7 (RNA-Seq report)	Both	1.81	98.19	0
DL2_S3 (RNA-Seq report)	Both	1.75	98.25	0
DL3_S2 (RNA-Seq report)	Both	1.47	98.53	0
AAV1_S12 (RNA-Seq report)	Both	1.69	98.31	0
AAV2_S9 (RNA-Seq report)	Both	1.92	98.08	0
AAV3_S6 (RNA-Seq report)	Both	1.70	98.30	0
PB1_S5 (RNA-Seq report)	Both	1.58	98.42	0
PB2_S10 (RNA-Seq report)	Both	1.31	98.69	0
PB3_S8 (RNA-Seq report)	Both	1.67	98.33	0
Minimum	-	1.17	98.08	0.00

Sample name	Strand specific setting	Forward reads mapped (%)	Reverse reads mapped (%)	Ignored reads (wrong strand) (%)
Median	-	1.68	98.32	0.00
Maximum	-	1.92	98.83	0.00
Mean	-	1.61	98.39	0.00
Standard deviation	-	0.21	0.21	0.00

Strand specific setting: >90% of reads were mapped in the same orientation. Consider re-running the tool with a strand specific setting ("Forward"/"Reverse").

1.5 Adapter read-through

No data available

1.6 Distribution of biotypes

The table is based on 12 samples.

Sample name	protein_coding	nonsense_mediated_ decay	processed_transcript	retained_intron
NL1_S1 (RNA-Seq report)	96.49	1.05	0.88	0.88
NL2_S11 (RNA-Seq report)	96.51	1.01	0.82	0.87
NL3_S4 (RNA-Seq report)	96.49	1.01	0.85	0.86
DL1_S7 (RNA-Seq report)	96.16	1.12	0.97	0.95
DL2_S3 (RNA-Seq report)	96.75	1.06	0.85	0.82
DL3_S2 (RNA-Seq report)	96.53	1.08	0.86	0.84
AAV1_S12 (RNA-Seq report)	96.59	1.04	0.85	0.73
AAV2_S9 (RNA-Seq report)	96.60	0.96	0.85	0.76
AAV3_S6 (RNA-Seq report)	96.12	0.96	0.83	0.79
PB1_S5 (RNA-Seq report)	96.46	0.89	0.86	0.93
PB2_S10 (RNA-Seq report)	96.71	0.94	0.80	0.76
PB3_S8 (RNA-Seq report)	96.39	1.05	0.88	0.94
Minimum	96.12	0.89	0.80	0.73
Median	96.50	1.03	0.85	0.85

Sample name	protein_coding	nonsense_mediated_decay	processed_transcript	retained_intron
Maximum	96.75	1.12	0.97	0.95
Mean	96.48	1.01	0.86	0.84
Standard deviation	0.19	0.06	0.04	0.07

Sample name	lincRNA	unprocessed_pseudogene	pseudogene	antisense
NL1_S1 (RNA-Seq report)	0.36	0.19	0.05	0.07
NL2_S11 (RNA-Seq report)	0.35	0.28	0.07	0.06
NL3_S4 (RNA-Seq report)	0.40	0.19	0.10	0.07
DL1_S7 (RNA-Seq report)	0.54	0.06	0.10	0.06
DL2_S3 (RNA-Seq report)	0.30	0.07	0.05	0.07
DL3_S2 (RNA-Seq report)	0.45	0.07	0.08	0.07
AAV1_S12 (RNA-Seq report)	0.40	0.12	0.06	0.06
AAV2_S9 (RNA-Seq report)	0.37	0.23	0.07	0.06
AAV3_S6 (RNA-Seq report)	0.77	0.19	0.20	0.06
PB1_S5 (RNA-Seq report)	0.46	0.19	0.10	0.06
PB2_S10 (RNA-Seq report)	0.45	0.16	0.09	0.05
PB3_S8 (RNA-Seq report)	0.40	0.17	0.05	0.07
Minimum	0.30	0.06	0.05	0.05
Median	0.40	0.18	0.07	0.06
Maximum	0.77	0.28	0.20	0.07
Mean	0.44	0.16	0.08	0.06
Standard deviation	0.12	0.07	0.04	0.00

Sample name	IG_C_gene	polymorphic_pseudogene	IG_V_gene	transcribed_unprocessed_pseudogene
NL1_S1 (RNA-Seq report)	0.00	0.01	0.00	0.00
NL2_S11 (RNA-Seq report)	0.01	0.01	0.00	0.00
NL3_S4 (RNA-Seq report)	0.02	0.01	0.00	0.00

Sample name	IG_C_gene	polymorphic_pseudogene	IG_V_gene	transcribed_unprocessed_pseudogene
DL1_S7 (RNA-Seq report)	0.00	0.02	0.00	0.00
DL2_S3 (RNA-Seq report)	0.00	0.01	0.00	0.00
DL3_S2 (RNA-Seq report)	0.00	0.01	0.00	0.00
AAV1_S12 (RNA-Seq report)	0.10	0.01	0.03	0.00
AAV2_S9 (RNA-Seq report)	0.05	0.01	0.02	0.00
AAV3_S6 (RNA-Seq report)	0.05	0.01	0.01	0.00
PB1_S5 (RNA-Seq report)	0.01	0.01	0.00	0.00
PB2_S10 (RNA-Seq report)	0.01	0.01	0.00	0.00
PB3_S8 (RNA-Seq report)	0.01	0.02	0.00	0.01
Minimum	0.00	0.01	0.00	0.00
Median	0.01	0.01	0.00	0.00
Maximum	0.10	0.02	0.03	0.01
Mean	0.02	0.01	0.01	0.00
Standard deviation	0.03	0.00	0.01	0.00

Sample name	processed_pseudogene	bidirectional_promoter_lncRNA	sense_intronic	non_stop_decay
NL1_S1 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL2_S11 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL3_S4 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL1_S7 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL2_S3 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL3_S2 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV1_S12 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV2_S9 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV3_S6 (RNA-Seq report)	0.00	0.00	0.00	0.00

Sample name	processed_pseudogene	bidirectional_promoter_lncRNA	sense_intronic	non_stop_decay
PB1_S5 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB2_S10 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB3_S8 (RNA-Seq report)	0.00	0.00	0.00	0.00
Minimum	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00
Maximum	0.00	0.00	0.00	0.00
Mean	0.00	0.00	0.00	0.00
Standard deviation	0.00	0.00	0.00	0.00

Sample name	unitary_pseudogene	TR_C_gene	TR_V_gene	IG_V_pseudogene
NL1_S1 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL2_S11 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL3_S4 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL1_S7 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL2_S3 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL3_S2 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV1_S12 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV2_S9 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV3_S6 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB1_S5 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB2_S10 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB3_S8 (RNA-Seq report)	0.00	0.00	0.00	0.00
Minimum	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00
Maximum	0.00	0.00	0.00	0.00
Mean	0.00	0.00	0.00	0.00

Sample name	unitary_pseudogene	TR_C_gene	TR_V_gene	IG_V_pseudogene
Standard deviation	0.00	0.00	0.00	0.00

Sample name	transcribed_processed_pseudogene	sense_overlapping	IG_LV_gene	TEC
NL1_S1 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL2_S11 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL3_S4 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL1_S7 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL2_S3 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL3_S2 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV1_S12 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV2_S9 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV3_S6 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB1_S5 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB2_S10 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB3_S8 (RNA-Seq report)	0.00	0.00	0.00	0.00
Minimum	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00
Maximum	0.00	0.00	0.00	0.00
Mean	0.00	0.00	0.00	0.00
Standard deviation	0.00	0.00	0.00	0.00

Sample name	translated_processed_pseudogene	transcribed_unitary_pseudogene	macro_lncRNA	TR_V_pseudogene
NL1_S1 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL2_S11 (RNA-Seq report)	0.00	0.00	0.00	0.00
NL3_S4 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL1_S7 (RNA-Seq report)	0.00	0.00	0.00	0.00
DL2_S3 (RNA-Seq report)	0.00	0.00	0.00	0.00

Sample name	translated_processed_pseudogene	transcribed_unitary_pseudogene	macro_lncRNA	TR_V_pseudogene
DL3_S2 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV1_S12 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV2_S9 (RNA-Seq report)	0.00	0.00	0.00	0.00
AAV3_S6 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB1_S5 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB2_S10 (RNA-Seq report)	0.00	0.00	0.00	0.00
PB3_S8 (RNA-Seq report)	0.00	0.00	0.00	0.00
Minimum	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00
Maximum	0.00	0.00	0.00	0.00
Mean	0.00	0.00	0.00	0.00
Standard deviation	0.00	0.00	0.00	0.00

Sample name	TR_J_gene	IG_J_gene	3prime_overlapping_ncRNA	IG_C_pseudogene
NL1_S1 (RNA-Seq report)	0.00	0.00	0.00	0
NL2_S11 (RNA-Seq report)	0.00	0.00	0.00	0
NL3_S4 (RNA-Seq report)	0.00	0.00	0.00	0
DL1_S7 (RNA-Seq report)	0.00	0.00	0.00	0
DL2_S3 (RNA-Seq report)	0.00	0.00	0.00	0
DL3_S2 (RNA-Seq report)	0.00	0.00	0.00	0
AAV1_S12 (RNA-Seq report)	0.00	0.00	0.00	0
AAV2_S9 (RNA-Seq report)	0.00	0.00	0.00	0
AAV3_S6 (RNA-Seq report)	0.00	0.00	0.00	0
PB1_S5 (RNA-Seq report)	0.00	0.00	0.00	0
PB2_S10 (RNA-Seq report)	0.00	0.00	0.00	0

Sample name	TR_J_gene	IG_J_gene	3prime_overlapping_ncRNA	IG_C_pseudogene
PB3_S8 (RNA-Seq report)	0.00	0.00	0.00	0
Minimum	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00
Maximum	0.00	0.00	0.00	0.00
Mean	0.00	0.00	0.00	0.00
Standard deviation	0.00	0.00	0.00	0.00

Sample name	TR_D_gene	IG_D_gene	IG_pseudogene
NL1_S1 (RNA-Seq report)	0	0	0
NL2_S11 (RNA-Seq report)	0	0	0
NL3_S4 (RNA-Seq report)	0	0	0
DL1_S7 (RNA-Seq report)	0	0	0
DL2_S3 (RNA-Seq report)	0	0	0
DL3_S2 (RNA-Seq report)	0	0	0
AAV1_S12 (RNA-Seq report)	0	0	0
AAV2_S9 (RNA-Seq report)	0	0	0
AAV3_S6 (RNA-Seq report)	0	0	0
PB1_S5 (RNA-Seq report)	0	0	0
PB2_S10 (RNA-Seq report)	0	0	0
PB3_S8 (RNA-Seq report)	0	0	0
Minimum	0.00	0.00	0.00
Median	0.00	0.00	0.00
Maximum	0.00	0.00	0.00
Mean	0.00	0.00	0.00
Standard deviation	0.00	0.00	0.00

1.7 Transcript length coverage

The table is based on 12 samples.

Sample name	Expected coverage bias	Difference between average 3’ and 5’ normalized counts	Reads mapping to transcripts that are longer than 10,000 bp (%)
NL1_S1 (RNA-Seq report)	Unbiased	8.37	2.22
NL2_S11 (RNA-Seq report)	Unbiased	8.48	2.03
NL3_S4 (RNA-Seq report)	Unbiased	10.86	2.85
DL1_S7 (RNA-Seq report)	Unbiased	11.34	2.44
DL2_S3 (RNA-Seq report)	Unbiased	13.64	2.40
DL3_S2 (RNA-Seq report)	Unbiased	12.31	3.28
AAV1_S12 (RNA-Seq report)	Unbiased	9.37	2.94
AAV2_S9 (RNA-Seq report)	Unbiased	8.87	2.01
AAV3_S6 (RNA-Seq report)	Unbiased	9.40	3.02
PB1_S5 (RNA-Seq report)	Unbiased	8.48	3.23
PB2_S10 (RNA-Seq report)	Unbiased	7.44	2.47
PB3_S8 (RNA-Seq report)	Unbiased	9.17	1.93
Minimum	-	7.44	1.93
Median	-	9.27	2.45
Maximum	-	13.64	3.28
Mean	-	9.81	2.57
Standard deviation	-	1.84	0.48