Bias Correction in RNAseq

Walter L. (Larry) Ruzzo

Computer Science and Engineering
Genome Sciences (Adjunct)
University of Washington
Fred Hutchinson Cancer Research Center (Joint Member)
Seattle, WA, USA
A new approach to bias correction in RNA-Seq

Daniel C. Jones,

Walter L. Ruzzo, Xinxia Peng and Michael G. Katze

Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350,

Department of Genome Sciences, University of Washington, Seattle, WA 98195-5065, Fred Hutchinson Cancer Research Center, Seattle, WA 98109 and Department of Microbiology, University of Washington, Seattle, WA

Associate Editor: Alex Bateman

ABSTRACT

Motivation: Quantification of sequence abundance in RNA-Seq experiments is often conflated by protocol-specific sequence bias. The exact sources of the bias are unknown, but may be influenced by factors such as cross-hybridization, RNA-Seq presents a new set of challenges. As microarray technology has brought with it technical challenges ranging from developing robust normalization to accounting for non-uniformity in nucleotide frequencies on five datasets (Table 1), position. Yet, this is not the case. As illustration, Figure 1 plots this quantification.

We evaluate our method on several datasets, and by multiple criteria, not rely on existing gene annotations, and model selection is performed automatically making it applicable with few assumptions.

These biases may adversely affect differential expression analysis, gene ontology enrichment tests, and mixtures, for example. The result is decreased accuracy in polymerase chain reaction amplification, or differing primer affinities or isoforms, and any test of differential expression hangs on the assumption that these biases are identical between replicates, an untrustworthy comparisons of relative abundance between genes.

We present a new method to measure and correct for many applications, such as

We demonstrate that it effectively decreases bias and increases uniformity. Additionally, we provide theoretical and empirical results showing that the method is unlikely to have any effect on unbiased quantification.

We provide parameter estimates for use in RNA-Seq experiments.

Availability: seqbias

 seqbias is available from http://bioconductor.org/R/Bioconductor package, available freely under the LGPL license by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/© The Author(s) 2012. Published by Oxford University Press.

Introduction

1 INTRODUCTION

Gene expression

The transcript is Poisson distributed with parameter $\lambda_i$, inhomogeneous Poisson process. The read count at position $i$ is then adjusted by the ratio of the foreground and background heptamer probabilities. Specifying two distributions over heptamers (i.e. foreground and background distributions) requires $2k\times2^k$ parameters, so while no gene annotations are hard-coded, chosen in advance using a hill-climbing algorithm variable-order Markov chains. The structure of these Markov chains are needed, and a number that increases exponentially with $k$, where $k$ is the length of the heptamer. The result is decreased accuracy in capturing the abundance of the transcript.

Another model, proposed by Hansen et al. (2010), directly models sequence probabilities by a linear model, in which read counts across a transcript follow an inhomogeneous Poisson process. The read count at position $i$ is then estimated by the model shows a moderate improvement over the Poisson linear model. Both models are fit to a number of abundant test genes, or MART (Friedman and Meulman, 2003). In their tests, the MART model shows a moderate improvement over the Poisson linear model. In their tests, the MART model shows a moderate improvement over the Poisson linear model.

In their tests, the MART model shows a moderate improvement over the Poisson linear model. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/© The Author(s) 2012. Published by Oxford University Press.

Results

We evaluate our method on several datasets, and by multiple criteria,

We demonstrate that it effectively decreases bias and increases uniformity. Additionally, we provide theoretical and empirical results showing that the method is unlikely to have any effect on unbiased quantification.
Example
What we expect: Uniform Sampling

Uniform sampling of 4000 “reads” across a 200 bp “exon.”
Average 20 ± 4.7 per position, min ≈ 9, max ≈ 33
i.e., as expected, we see ≈ μ ± 3σ in 200 samples
What we get: *highly non-uniform coverage*

E.g., assuming uniform, the 8 peaks above 100 are $\geq +10\sigma$ above mean.
What we get: *highly* non-uniform coverage

The Good News: we can (partially) correct the bias

*not perfect, but better: 38% reduction in LLR of uniform model; hugely more likely*
Bias is sequence-dependent

and platform/sample-dependent

Fitting a model of the sequence surrounding read starts lets us predict which positions have more reads.
To estimate sequencing bias, we train a Bayesian network in which each show the results of this method applied to the 3 is evaluated at each position within a locus, predicting bias. The predicted and background sequence probabilities. They differ primarily in how these 2.2 Estimation (are sampled from the regions surrounding the starts of mapped reads; foreground training sequence is paired with a background sequence taken from a nearby position that is likely to have similar abundance and general risk of incorrectly adjusting for biological sequence bias, rather than technical background sequences are sampled uniformly from the genome. we run the Notably, exonic DNA tends to be more GC-rich than intergenic DNA. If model.

\[
\log_{10} \text{bias} = 1.14 \text{ and } 0.63, \text{ suggesting that most adjustments are not large.}
\]

An overview of the approach taken: (1.14 and 0.63, suggesting that most adjustments are not large. (a) sample foreground sequences

(b) sample background sequences

(c) train Bayesian network

(d) predict bias

(e) adjust read counts

\[
R^2 = 0.38
\]
Want a probability distribution over k-mers, $k \approx 40$

Some obvious choices

- Full joint distribution: $4^k - 1$ parameters

- PWM (0-th order Markov): $(4 - 1) \cdot k$ parameters

Something intermediate

- Directed Bayes network
Form of the models: Directed Bayes nets

Wetterbom (282 parameters)

One “node” per nucleotide, ±20 bp of read start
- Filled node means that position is biased
- Arrow $i \rightarrow j$ means letter at position $i$ modifies bias at $j$
- For both, numeric parameters say how much

How—optimize:

$$\ell = \sum_{i=1}^{n} \log \Pr[x_i | s_i] = \sum_{i=1}^{n} \log \frac{\Pr[s_i | x_i] \Pr[x_i]}{\sum_{x \in \{0,1\}} \Pr[s_i | x] \Pr[x]}$$
unpublished dataset of our own (Table 1).

2008; Trapnell correction (Section 3.3).

real-time PCR (qRT-PCR) based estimates for the same genes,

compare RNA-Seq-based estimate of gene expression to quantitative

read coverage (Poisson regression in Section 3.2). Additionally, we

Leibler divergence in Section 3.1) and increase in uniformity of

that could be applied genome-wide. To this end, we used cross-

often does not hold in typical RNA-Seq datasets. Although the

our evaluation strategy relies on testing three assumptions we make

indicate positions that were not included in the model, being deemed uninformative, given the other positions and edges. The number of parameters needed

(1) Positional nucleotide frequencies (as in Fig. 1), measured from

max

(2) Read counts across a single exon should follow,\text{max}

approximately, a Poisson process.

reads within exons, should not differ greatly from frequencies

measured by sampling uniformly within the same exons.

reads from chromosomes 1–8, considering the initial read position

and 20 nt upstream and downstream.

All datasets were mapped using Bowtie (Langmead

et al., 2008; Wetterbom

et al., 2010), as well as an

Genominator

Hubbard

and Mortazavi

at University of Washington on October 9, 2012

Figure 3 shows examples of the structure learned when this procedure is

reimplemented the model and trained it on all of the reads aligned

we were unable to train with the volume of data we wished, so we

heptamer frequencies used by their model. The training procedure

training by prohibiting very distant dependencies (>10, by default) or very high in-degrees (> 4, by default).

validation tests (i.e. methods were trained and tested on disjoint

number

Training by prohibiting very distant dependencies (>10, by default) or very high in-degrees (> 4, by default).

remaining chromosomes. In particular, for each reference sequence,

of the genome from which the reads were mapped (including

rheMac2 and panTro2 genome assemblies obtained from the UCSC

sequences was at least the length of the exon.

evaluation, we drew a set of long, highly expressed exons from the

ourselves to those at least 100 nt long.

 succeeds in the evaluation, we trained on the 100 000 randomly selected

exons from chromosomes 1–8. MART models were trained considering the initial read position

100 most abundant genes. We used 1000 exons

Each method was trained on data taken from chromosomes 1–8

each method is listed in Section 3 in Supplementary Material, and

datasets, using 100 000 reads from each.

validation procedures unique to each. The total number of reads used to train

mseq

−

423

• Not just initial hexamer

• Span ≥ 19

• All include negative positions
Formally...

A reasonable definition of unbiasedness:

\[ \Pr(\text{read at i}) = \Pr(\text{read at i}|\text{sequence at i}) \]

From Bayes...

\[ \Pr(\text{read at i}|\text{sequence at i}) = \frac{\Pr(\text{sequence at i}|\text{read at i}) \Pr(\text{read at i})}{\Pr(\text{sequence at i})} \]

So we might define bias as

\[ \text{bias at position } i = \frac{\Pr(\text{sequence at i}|\text{read at i})}{\Pr(\text{sequence at i})} \]
Conditional Log-Likelihood

Find a graph that maximizes conditional log-likelihood.

\[
\text{CLL} = \sum_{i=1}^{n} \log \Pr(x_i | s_i)
\]

We need to penalize for model complexity as well.

\[
\text{CLL}' = 2 \sum_{i=1}^{n} \log \Pr(x_i | s_i) - m \log n
\]
Yet, given the high coverage of the exons being tested, if duplicate remaining cases are plotted in Section 4 in Supplementary Material.

When computing the KL divergence, there is a risk of the measure frequencies will result in a larger KL divergence. To quantify the non-uniformity observed in these plots, we use the following method: all reads contained within the exon being tested reads are excluded, it may not capture the full effect of bias frequencies will result in a larger KL divergence.

We compute the KL divergence to be low. We compute the half only once, ignoring duplicates. For comparison, we then fit a model in which the rate is

\[ \lambda_{ij} = \frac{f_x}{\sum_j f_{x,j}} \]

for a \( k \)-mer distribution \( f_x \), \( f_{x,j} \) for \( j \) and \( k \)-mers. This can be thought of as a

\[ D_{KL}(f_{x,j}) = \sum f_x \log \left( \frac{f_x}{f_{x,j}} \right) \]

for \( i \) and

\[ D_{KL}(f_x) = \frac{1}{2} \sum f_x \log \left( \frac{f_x}{f_x'} \right) \]

where \( f_x' \) is the background frequency of a

\[ f_x' = \frac{\lambda_{ij}}{\sum_j \lambda_{ij}} \]

\[ f_{x,j} = \frac{\lambda_{ij}}{\sum_i \lambda_{ij}} \]

The KL divergence compares the frequency of the observed

\[ f_x \]

and

\[ f_x' \]

The assumption of positional read counts following a Poisson distribution

\[ \lambda_{ij} = \frac{f_x}{\sum_j f_{x,j}} \]

We repeated the procedure for

\[ k = 1 \]

and

\[ k = 4 \]

values <0, indicating that the model has worse fit than the null model (a large improvement, corresponding to, for example, the null model, the Poisson rate is fixed for each test exon. That is, for

\[ \lambda \]

being fit. For comparison, we then fit a model in which the rate is

\[ \lambda \]

for bias remains a principled and easily interpreted criterion. This increase in uniformity is illustrated in Figure 2.

\[ L \]

\[ L \]

also proportional to the predicted bias coefficients:
In a comparison, the MART method performs equally.

Improvement in performance over other methods in all but one listed in Table 2. Our method shows a highly statistically significant dots marking more extreme values.

In the inter-quartile range (i.e., the span between the 25% and 75% quantiles), and 5 times the 25%, 50%, and 75% quantiles, with whiskers extending to 1.5 times the inter-quartile range.

For each of the 1000 test exons, we compute McFadden's pseudo-

The Kullback-Leibler divergence compares the frequency of expected under the assumption of uniform sampling.

Marked with asterisks are methods over which the B decreases. Marked with two asterisks are methods over which the B increases as uniformity is increased, and negative when uniformity is decreased.

The likelihood under the bias corrected model. The statistic is p.

This evaluation does not rely on an assumption that qRT-PCR is necessarily more accurate than RNA-Seq based quantification, only data shows a pattern of bias similar to that seen in the other samples.

We used sequencing data previously published by Au et al. (2010) to evaluate the effect bias correction has on correlation to measurements made by TaqMan RT-PCR, made available by the the Microarray Quality Control project (Shi et al., 2006). The RNA-Seq data shows a pattern of bias similar to that seen in the other samples.

Table 2. The goodness of fit statistic measures increased uniformity in read coverage, after correcting for bias. Here the median across test exons is listed for each method and sample.

The statistic measures increased uniformity in read coverage, after correcting for bias. Here the median across test exons is listed for each method and sample.

The coefficient of determination.
The RNA-Seq data shows a pattern of bias similar to that seen in Table 2. Considered, we counted reads overlapping each gene, defining the same way as the RNA-Seq data. Seq-based quantification, only that qRT-PCR is not biased in the assumption that qRT-PCR is necessarily more accurate than RNA-Seq. In Supplementary Material.) This evaluation does not rely on an in the other samples sequenced on an Illumina platform (Section 6 in Supplementary Material.) The same is true of the Mortazavi data. In all our evaluation, no special tuning of need only be greater than zero for an adequate model to be trained.

### Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>7mer</th>
<th>MART</th>
<th>GLM</th>
<th>BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>r^2</td>
<td>0.163</td>
<td>0.224</td>
<td>0.157</td>
<td>0.243</td>
</tr>
<tr>
<td>*p &lt; 10^-23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fractional improvement in log-likelihood under uniform model across 1000 exons (R^2=1-L'/L)

---

**Result – Increased Uniformity**

**Wetterbom**

**Katze**

**BN**

**MART**

**GLM**

**7mer**

hypothesis test: “Is BN better than X?” (1-sided Wilcoxon signed-rank test)

* = p-value < 10^-23

---

To evaluate the efficacy of each of the bias correction methods, we estimated the statistical significance of the improvement in correlation using the hypothesis test:

\[ R^2 = 1 - L'/L \]

where \( R^2 \) is the correlation coefficient, \( L' \) is the log-likelihood of the corrected model, and \( L \) is the log-likelihood of the uncorrected model. This test is one-sided, comparing BN against the other methods. The highest value in each row is marked with a single asterisk (*), and those marked with two asterisks (***) achieved a higher correlation than the BN method. The box plots show the 25th, 50th, and 75th percentiles, with whiskers extending to 1.5 times the interquartile range. Fractional improvement in log-likelihood under uniform model across 1000 exons (R^2=1-L'/L).

* = p-value < 10^-23
Fig. 4. The Kullback-Leibler divergence compares the frequency of \( k \)-mers (here, for \( k = 1 \) and \( k = 4 \)) surrounding the start of an aligned read to the frequencies expected under the assumption of uniform sampling from exons. A large divergence indicates a bias. Plotted here is the divergence from unadjusted read counts as well as after adjusting read counts using each method.

Fig. 5. For each of the 1000 test exons, we compute McFadden's pseudo-coefficient of determination \( R^2 \), equivalent to the improvement in log-likelihood under the bias corrected model. The statistic is positive, and increases as uniformity is increased, and negative when uniformity is decreased. Marked with asterisks are methods over which the BN approach showed a statistically significant improvement when applied to the same data, according to a one-sided Wilcoxon signed-rank test. Above each of those marked, we observed p-values less than $10^{-23}$. Boxes are plotted to mark the 25%, 50%, and 75% quantiles, with whiskers extending to 1.5 times the inter-quartile range (i.e., the span between the 25% and 75% quantiles), and dots marking more extreme values.

Table 2. The median \( R^2 \) goodness of fit statistic across test exons.

<table>
<thead>
<tr>
<th></th>
<th>BN</th>
<th>MART</th>
<th>GLM</th>
<th>7mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wetterbom</td>
<td>0.174</td>
<td>0.016</td>
<td>0.066</td>
<td>-0.079</td>
</tr>
<tr>
<td>Katze</td>
<td>0.280</td>
<td>0.243</td>
<td>0.158</td>
<td>0.033</td>
</tr>
<tr>
<td>Bullard</td>
<td>0.267</td>
<td>0.163</td>
<td>0.224</td>
<td>0.157</td>
</tr>
<tr>
<td>Mortazavi</td>
<td>0.240</td>
<td>0.210</td>
<td>0.197</td>
<td>0.091</td>
</tr>
<tr>
<td>Trapnell</td>
<td>0.289</td>
<td>0.289</td>
<td>0.248</td>
<td>0.138</td>
</tr>
</tbody>
</table>

The \( R^2 \) statistic measures increased uniformity in read coverage, after correcting for bias. Here the median \( R^2 \) across the test exons is listed for each method and sample. A higher \( R^2 \) indicates a better fit. The highest value in each row is highlighted in bold.

3.3 qRT-PCR Correlation

We used sequencing data previously published by Au et al. (2010) to evaluate the effect bias correction has on correlation to measurements made by TaqMan RT-PCR, made available by the the Microarray Quality Control project (Shi et al., 2006). The RNA-Seq data shows a pattern of bias similar to that seen in the other samples sequenced on an Illumina platform (see Supplementary Section 6). This evaluation does not rely on an assumption that qRT-PCR is necessarily more accurate than RNA-Seq based quantification, only that qRT-PCR is not biased in the same way as the RNA-Seq data.

To evaluate the efficacy of each of the bias correction methods considered, we counted reads overlapping each gene, defining the gene by the union of every transcript in release 60 of the Ensembl gene annotations. Counts were then normalized by dividing by the length of these genes. We then removed any genes with a read count less than 10, or that did not correspond to a unique TaqMan probe. Each method was trained in a manner identical to that used in the analysis of section 3.1 and 3.2, but without restricting the training data to the first eight chromosomes. After adjusting read counts according to the predicted sequence bias, we computed the Pearson correlation coefficient \( r \) between log read counts and log TaqMan expression values, which are averaged across three replicates. The \( r \) is listed in Table 2.
“First, do no harm”

Theorem:

The probability of “false bias discovery,” i.e., of learning a non-empty model from \( n \) reads sampled from unbiased data is less than

\[
1 - \left( \Pr(X < 3 \log n) \right)^{2h}
\]

where \( h \) = number of nucleotides in the model and \( X \) is a random variable that (asymptotically in \( n \)) is \( \chi^2 \) with 3 degrees of freedom. (\( E[X] = 3 \))
how different are two distributions?

Given: r-sided die, with probs p₁...pₙ of each face. Roll it n=10,000 times; observed frequencies = q₁, ..., qₙ, (the MLEs for the unknown qᵢ’s). How close is pᵢ to qᵢ?

Kullback-Leibler divergence, also known as relative entropy, of Q with respect to P is defined as

$$H(Q | | P) = \sum q_i \ln \frac{q_i}{p_i}$$

where qᵢ (pᵢ) is the probability of observing the iᵗʰ event according to the distribution Q (resp., P), and the summation is taken over all events in the sample space (e.g., all k-mers). In some sense, this is a measure of the dissimilarity between the distributions: if pᵢ ≈ qᵢ everywhere, their log ratios will be near zero and H will be small; as qᵢ and pᵢ diverge, their log ratios will deviate from zero and H will increase.

Fancy name, simple idea: H(Q | | P) is just the expected per-sample contribution to log-likelihood ratio test for “was X sampled from H₀: P vs H₁: Q?”

So, assuming the null hypothesis is false, in order for it to be rejected with say, 1000 : 1 odds, one should choose m to be inversely proportional to H(Q | | P):

$$mH(Q | | P) \geq \ln 1000$$

$$m \geq \frac{\ln 1000}{H(Q | | P)}$$
Continuing the notation above, suppose \( P \) as an unknown distribution with parameters \( p_1, \ldots, p_r \), \( \sum p_i = 1 \) where \( r \) is the number of points in the sample space (e.g. \( r = 4^k \) in the case of \( k \)-mers). Given a random sample \( X_1, X_2, \ldots, X_r \) of size \( n = \sum_i X_i \) from \( P \), it is well known that the maximum likelihood estimators for the parameters are \( q_i = \frac{X_i}{n} \approx p_i \). How good an estimate for \( P \) is this distribution \( Q \)? The estimators are unbiased:

\[
E[q_i] = E\left[ \frac{X_i}{n} \right] = E\left[ \frac{X_i}{n} \right] = \frac{n p_i}{n} = p_i
\]

and the standard deviation of each estimate is proportional to \( 1/\sqrt{n} \), so these estimates are increasingly accurate as the sample size increases. A more quantitative assessment of the accuracy of the estimator is obtained by evaluating the KL divergence:

\[
H(Q||P) = \sum_{i=1}^{r} q_i \ln \frac{q_i}{p_i} = \sum_{i=1}^{r} q_i \ln \left( 1 + \frac{q_i - p_i}{p_i} \right)
\]
Using the first two terms of the Taylor series for \( \ln(1 + x) \), this is

\[
H(Q \| P) \approx \sum_{i=1}^{r} q_i \left( \frac{q_i - p_i}{p_i} - \frac{1}{2} \left( \frac{q_i - p_i}{p_i} \right)^2 \right) \\
= \sum_{i=1}^{r} q_i \left( \frac{q_i - p_i}{p_i} - \frac{q_i^2}{2p_i} \right)
\]

Since \( \sum_{i=1}^{r} q_i = \sum_{i=1}^{r} p_i = 1 \), \( \sum_{i=1}^{r} p_i \frac{q_i - p_i}{p_i} = 0 \), so

\[
H(Q \| P) \approx \sum_{i=1}^{r} q_i \left( \frac{q_i - p_i}{p_i} - \frac{q_i}{2p_i} \right) \\
= \sum_{i=1}^{r} \frac{(q_i - p_i)^2}{p_i} \left( 1 - \frac{q_i}{2p_i} \right) \\
\approx \frac{1}{2} \sum_{i=1}^{r} \frac{(q_i - p_i)^2}{p_i}
\]

since \( q_i \approx p_i \). Multiplying by \( n^2/n^2 \) we have,

\[
H(Q \| P) \approx \frac{1}{2n} \sum_{i=1}^{r} \frac{(nq_i - np_i)^2}{np_i} \\
= \frac{1}{2n} \sum_{i=1}^{r} \frac{(X_i - E[X_i])^2}{E[X_i]}
\]
... and after a modicum of algebra:

\[ E[H(Q||P)] \approx \frac{r - 1}{2n} \]

... which empirically is a good approximation:

LLR of error rises with number of parameters \( r \), declines with size of training set \( n \)
What is the chance that we will learn an incorrect model? E.g., learn a biased model from unbiased input?

Wetterbom
(282 parameters)

Bullard
(696 parameters)

How does the amount of training data effect accuracy of the resulting model?
Probability of falsely inferring “bias” from an unbiased sample declines rapidly with size of training set (provably) ...

If > 10,000 reads are used, the probability of a non-empty model < 0.0004

Figure 8: Median $R^2$ is plotted against training set size. Each point is additionally labeled with the run time of the training procedure.

... while accuracy and runtime rise (empirically)

Training time: 10-50,000 reads in minutes; $10^5$ reads in an hour
Possible objection to the approach:

Typical expts compare gene A in sample 1 to itself in sample 2. Gene A’s sequence is unchanged, “so the bias is the same” & correction is useless/dangerous.

Responses:

Bias is *sample-dependent*, to an unknown degree.

*SNPs and/or alternative splicing* might have a big effect, if samples 1 & 2 are from different individuals and/or engender changes in isoform usage.

Some experiments are *not “typical,”* e.g., imprinting, allele specific expression, xenograft studies.

Strong control of “false bias discovery” ⇒ *little risk*.

In Progress: Isolator

Soon to be the world’s best isoform quantitation tool.
seqbias

Estimation of per-position bias in high-throughput sequencing data

Biocductor version: Release (2.12)

This package implements a model of per-position bias in high-throughput sequencing data. The model was estimated using a simple Bayesian network, the structure of which is illustrated above. The model is fitted to a set of simulated reads from the ENCODE dataset. The results are consistent with the observed biases in the data.

Author: Daniel Jones <dcjones at cs.washington.edu>
Maintainer: Daniel Jones <dcjones at cs.washington.edu>

To install this package, start R and enter:

source("http://bioconductor.org/biocLite("seqbias")

To cite this package in a publication:

citation("seqbias")

Document: Assessing and Adjusting for Bias in High-throughput Sequencing Data

Availability


Download stats for Software package seqbias

This page was generated on 2013-09-02 07:28:58 -0700 (Mon, 02 Sep 2013).

seqbias home page: release version, devel version.

<table>
<thead>
<tr>
<th>Month</th>
<th>Nb of distinct IPs</th>
<th>Nb of downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct/12</td>
<td>131</td>
<td>180</td>
</tr>
<tr>
<td>Nov/12</td>
<td>150</td>
<td>204</td>
</tr>
<tr>
<td>Dec/12</td>
<td>123</td>
<td>176</td>
</tr>
<tr>
<td>Jan/13</td>
<td>119</td>
<td>166</td>
</tr>
<tr>
<td>Feb/13</td>
<td>98</td>
<td>129</td>
</tr>
<tr>
<td>Mar/13</td>
<td>146</td>
<td>282</td>
</tr>
<tr>
<td>Apr/13</td>
<td>187</td>
<td>250</td>
</tr>
<tr>
<td>May/13</td>
<td>217</td>
<td>333</td>
</tr>
<tr>
<td>Jun/13</td>
<td>200</td>
<td>293</td>
</tr>
<tr>
<td>Jul/13</td>
<td>142</td>
<td>205</td>
</tr>
<tr>
<td>Aug/13</td>
<td>164</td>
<td>248</td>
</tr>
<tr>
<td>Sep/13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>All months</td>
<td>1321</td>
<td>2498</td>
</tr>
</tbody>
</table>