SOFTWARE TOOL ARTICLE

Restriction site-associated DNA from Python-implemented Digestion Simulations (RApyDS): a companion tool for RAD sequencing experimental design [version 1; peer review: awaiting peer review]

Kristianne Arielle Gabriel¹, Maria Rejane Nepacina ^{1,2}, Francis Tablizo¹, Carlo Lapid¹, Mark Lenczner Mendoza¹, Daniella Jean Pamulaklakin^{1,3}, Jobeth Domingo ^{1,4}, Marc Jermaine Pontiveros^{1,5}, Hannah Mae Magno^{1,6}, Joshua Gregor Dizon¹, El King Morado¹, Jan Michael Yap¹

¹Core Facility for Bioinformatics, Philippine Genome Center, University of the Philippines, Diliman, Quezon City, 1101, Philippines

²Science Education Institute, Department of Science and Technology, Bicutan, Taguig City, 1631, Philippines

³Institute of Weed Science, Entomology and Plant Pathology, College of Agriculture and Food Sciences, University of the Philippines - Los Banos, Laguna, 4031, Philippines

⁴Division of Biological Sciences, College of Arts and Sciences, University of the Philippines Visayas, Miagao, Iloilo, 5023, Philippines

⁵Department of Physical Sciences and Mathematics, College of Arts and Sciences, University of the Philippines Manila, Manila, 1004, Philippines

⁶Department of Computer Science, University of the Philippines Diliman, Quezon City, 1101, Philippines

V1 First published: 07 May 2021, 10:360
<https://doi.org/10.12688/f1000research.52141.1>
Latest published: 07 May 2021, 10:360
<https://doi.org/10.12688/f1000research.52141.1>

Open Peer Review

Approval Status AWAITING PEER REVIEW

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Reduced representation sequencing is a practical approach for obtaining genetic variations from a random subsample of the genome. RADseq (Restriction Site-Associated DNA Sequencing), as one of the more popular reduced representation approaches, is currently being used in a wide array of applications including marker development, phylogenetics, and population genomics. A crucial step in designing a RADseq experiment is the selection of one or a pair of restriction enzymes (RE) that will result in sufficient density of loci to meet the objectives of the study, which is not straightforward because of difficulties in obtaining a standard set of REs that can generally be applied to RADseq experimental designs. Here we present RApyDS, a simulation tool that provides users with evaluation metrics to aid in choosing suitable REs based on their target RADseq design. RApyDS can perform simulations for single- or double-digest RADseq, preferably with a supplied reference genome. The tool outputs an overview page, electrophoresis visualization, mapping of restriction cut sites, and RAD loci density across the genome. If supplied with an annotation file, the program can also output evaluation metrics for a

specified genomic feature. The tool is currently available at <https://github.com/pgcbioinfo/rapyds>.

Keywords

RApyDS, Restriction Site-Associated DNA Sequencing, RADseq, simulation, companion tool, experimental design

This article is included in the **Bioinformatics** gateway.

This article is included in the **Cell & Molecular Biology** gateway.

This article is included in the **Python** collection.

Corresponding author: Jan Michael Yap (jcyap@up.edu.ph)

Author roles: **Gabriel KA:** Conceptualization, Methodology, Software, Writing – Review & Editing; **Nepacina MR:** Conceptualization, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Tablizo F:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Lapid C:** Conceptualization, Writing – Review & Editing; **Mendoza ML:** Methodology, Software; **Pamulaklakin DJ:** Methodology, Software; **Domingo J:** Investigation, Validation; **Pontiveros MJ:** Investigation, Validation; **Magno HM:** Investigation, Validation; **Dizon JG:** Conceptualization, Software, Writing – Review & Editing; **Morado EK:** Conceptualization, Writing – Review & Editing; **Yap JM:** Conceptualization, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was partially funded by the University of the Philippines – Philippine Genome Center (UP-PGC) and the Philippine Council for Industry, Energy and Emerging Technology Research and Development of the Department of Science and Technology (DOST-PCIEERD) through the Increased Local Bioinformatics Utility and Expertise (IMBUE) project (DOST-PCIEERD Funding No. 2019-101) granted to JMC Yap.

Copyright: © 2021 Gabriel KA *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gabriel KA, Nepacina MR, Tablizo F *et al.* **Restriction site-associated DNA from Python-implemented Digestion Simulations (RApyDS): a companion tool for RAD sequencing experimental design [version 1; peer review: awaiting peer review]** F1000Research 2021, **10**:360 <https://doi.org/10.12688/f1000research.52141.1>

First published: 07 May 2021, **10**:360 <https://doi.org/10.12688/f1000research.52141.1>

Introduction

Next-generation sequencing (NGS) technologies provide an avenue to obtain a broader view of the genetic landscape of individual species, substantially impacting various applications such as molecular marker discovery (Shendure & Ji, 2008). However, not all approaches and studies require sequencing of complete genomes, which may be cost-prohibitive and more difficult to analyze (Etter *et al.*, 2011; Meger *et al.*, 2019). In dealing with whole-genome resequencing studies, reduced representation genomic libraries (RRLs) are alternatively used to randomly subsample the genome for the discovery of genetic markers and variations. Instead of resequencing at a lower depth of coverage to reduce the sequencing cost, RRLs can increase the effective coverage depth but only for certain regions of the genome. Amongst RRLs, Restriction Site-Associated DNA Sequencing (RADseq) has gained a lot of attention, as this approach utilizes the power of high-throughput sequencing to efficiently discover polymorphic markers on organisms even in the absence of a reference genome (Etter *et al.*, 2011).

RADseq, which utilizes restriction enzymes to sample a subset of the genome that is flanked by restriction sites, is commonly used today for SNP discovery and genotyping (Davey *et al.*, 2013). RADseq protocols may vary according to the number of restriction enzymes used and the frequency of cut sites across the length of the DNA. In its original formulation, RADseq relies on the fragments produced from the cut sites of a single restriction enzyme; whereas other protocols such as double-digest RADseq (ddRAD) produce fragments with known distances from the cut-sites of two different restriction enzymes (Andrews *et al.*, 2016). Briefly, the process begins with the digestion of high-quality genomic DNA by the chosen restriction enzyme. Molecular identifiers and adapters are ligated to the sticky ends of the fragments. Samples are then pooled and randomly sheared. Another set of adapters are attached to the sheared fragments to ensure proper amplification. Finally, size selection is performed for sequence-ready fragments. Random shearing is eliminated in the ddRAD protocol as it uses another restriction enzyme for a second round of digestion instead (Davey & Blaxter, 2010).

Restriction enzymes are the key elements in RADseq. These are essential in determining the genomic regions to be sequenced, since the patterns of genetic variations to be observed will occur adjacent to the corresponding restriction sites (Andrews *et al.*, 2016). The choice of which enzyme to use thus determines the number of markers that can be obtained and assayed, which in turn determines the amount of sequencing required, the number of samples, the cost, and ultimately the success of the project (Herrera *et al.*, 2015). The lack of prior estimates on the number of restriction sites, fragments that can be produced, density of RAD loci that can be sequenced, or other projected effects of using one or a combination of restriction enzymes are some of the common impediments in successfully designing and performing reduced representation library preparations for whole-genome resequencing. It is therefore of utmost importance to have a prior idea on the distribution and

frequency of cut-sites and fragments that will be produced by a given restriction enzyme.

To address the aforementioned impediments, we developed RApYDS (Restriction Site-Associated DNA from Python-implemented Digestion Simulation), a tool that aims to provide evaluation metrics to aid researchers in choosing the best set of restriction enzymes for a RADseq study. It is a Python-based tool that simulates digestion of a given reference genome in silico either by original RADseq or ddRAD protocols. The simulation tool generates estimates for the following: (i) the number of restriction sites, (ii) the number of fragments after digestion, (iii) the number of fragments after shearing and size selection (RAD loci), (iv) the number of RAD loci within repeat regions, (v) the cut-site distribution, and (vi) the projected RAD loci density across the genome. A gel electrophoresis view of the restriction fragments is also shown to simulate the common laboratory method used for initial evaluation of restriction digestion. When supplied with an annotation file for the reference genome, the tool will also provide the number of RAD loci observed within annotated regions. All of this information, taken together, will enable the user to better determine the top restriction enzyme candidates and RADseq experimental design to implement for a particular RADseq or resequencing study.

Methods

Implementation

Figure 1 shows an overview of the implementation workflow of the RApYDS tool. Two distinct RADseq protocols are implemented in RApYDS for simulation: the original RADseq protocol (Baird *et al.*, 2008), and the double-digest RADseq or ddRAD (Peterson *et al.*, 2012). Global splitting was implemented to represent the digestion of DNA sequence based on the selected restriction enzymes; one RE for original RAD and two REs for ddRAD. In order to obtain simulation metrics based on the mapping of RAD loci to the reference genomic sequence, we used the alignment tool BWA (BWA, RRID:SCR_010910) (Li & Durbin, 2009). At present, the primary output of RApYDS is an interactive HTML file that can be viewed properly using the Mozilla Firefox browser.

Original RADseq Simulation. In this protocol, we simulate the original RADseq workflow by initially segmenting the reference sequence based on the presence of restriction enzyme cut sites. The procedure therefore assumes a complete digestion of the intact genome. The resulting fragments are then further sheared and size-selected according to a user-defined insert size range. This process results into three fragment types: (1) those that are deemed too small after shearing and size selection, (2) those that are within the insert size range but do not contain a sticky end (remnant of the restriction cut site), and (3) those that are within the defined insert size range and contains both sticky and sheared ends. Type 1 fragments are filtered out by the program to simulate the size selection process in sequencing library preparation. Type 2 fragments, on the other hand, are also filtered out by the program because the sticky end resulting from the restriction enzyme activity is

Figure 1. Summary of RApYDS workflow.

needed for the attachment of the P1 sequencing adapter. Thus, only fragments containing a restriction enzyme cut site in one of the ends and whose lengths are within the insert size range (Type 3) will be considered as putative RAD loci for downstream analysis.

Double-digest RADseq Simulation. For the ddRAD protocol, the process similarly begins with the segmentation of the genome based on the positions of the first restriction enzyme cut sites. However, instead of implementing a shearing step afterwards, a second round of digestion based on the cut sites of the second restriction enzyme is done. The resulting fragments after the two rounds of digestion are then size selected according to the user-defined insert size range. This procedure will generate the following fragment types: (1) those that are shorter than the minimum insert size, (2) those that are within the insert size range but contain sticky ends from only a single enzyme, and (3) those that are within the insert size range and contain sticky ends from both of the enzymes used. Again, to simulate size selection during library preparation, Type 1

fragments are filtered out in the process. Type 2 fragments will also be filtered out because the sticky ends from both enzymes are necessary for the proper attachment of the sequencing adaptor pairs. At the end of this step, only the Type 3 fragments will remain as putative RAD loci.

Operation

RApyDS software runs on a Linux OS computing machine with at least Python 3 installed and a minimum system requirement of 500 MB free memory (RAM) and a single CPU. Resource consumption will depend on the size of the input genome. Installation of the Python module dependencies can be done using pip. A complete guide on installation, usage, software requirements, and specifications can be accessed at <https://github.com/pgcbioinfo/rapyds>.

Use cases

Program inputs

The program requires a genome or a DNA sequence in FASTA format as input. Whenever a corresponding annotation in

GFF/GTF format is available, this can also be supplied together with the sequence to enable counting of RAD loci within target regions. In cases where a reference genome is not available, the tool can randomly generate a hypothetical reference sequence based on a user-defined GC content and sequence length using the algorithm implemented in SimRAD (Lepais & Weir, 2014). By default, the tool will evaluate a total of 56 commonly used restriction enzymes which recognition sequences were acquired from New England Biolabs Inc. . The users may also select any subset of these enzymes for evaluation and are given the option to add new restriction enzyme information if these are not yet included in the current list.

Outputs and implementation features

A successful RApYDS run will generate a compressed ZIP file as final output. This archived file contains simulation reports in hypertext markup language (HTML) format, all of which can be accessed interactively through the main file *index.html*. The landing page (Overview tab) shows a table containing various simulation metrics for each of the restriction enzymes tested (Figure 2). These metrics are described in more detail in Table 1. Briefly, the table contains information on the number of predicted RAD loci, the breadth of reference sequence

covered by these loci, single-copy and repetitive loci, and loci within annotated regions if an annotation file has been supplied. To further facilitate assessment of the results, the table can be sorted in increasing or decreasing order based on the values from a particular column by clicking on the column header.

We note here that the number of single-copy and repetitive RAD loci is based on alignment results. In certain cases, the alignment step may take a substantial amount of time to complete. Thus, we included a parameter flag to skip the alignment step if the user prefers to speed up the simulation process at the expense of repeat information. Table 2 shows the average processing times for the following simulations: (i) with alignment and reference indexing, (ii) with alignment only wherein an indexed reference is supplied, and (iii) without alignment. A reduction in processing time of around two to five times can be obtained when running RApYDS using the latter options.

In addition to the overview metrics, graphical views of the results in various forms are also made available to the users. The first is a virtual gel electrophoresis output based on the obtained fragments after digestion simulations (Figure 3). For this

RApyDS Report Overview Electrophoresis Cut Site Distribution									
NC_003279.8									
NC_003280.10									
NC_003281.10									
NC_003282.8									
NC_003283.11									
NC_003284.9									
NC_001328.1									
NC_003279.8 <i>Caenorhabditis elegans</i> chromosome I									
Restriction Enzyme	Fragments after digestion	Fragments after size selection (RAD loci)	Percent breadth of coverage	Single-copy RAD loci	Repetitive RAD loci	Repeat regions with RAD loci	RAD loci in specified feature	Specified feature with RAD loci	Percent specified feature with RAD loci
AatII	2691	4382	8.722	4311	71	148	2827	1024	25.321
Acc65I	1074	1922	3.826	1879	43	71	1219	539	13.328
AccI	6566	9448	18.805	9239	209	518	5761	1702	42.087
Acil	21595	16082	32.009	15815	267	856	9745	2180	53.907
AcII	3782	6246	12.432	6126	120	283	3633	1211	29.946
AfIII	417	790	1.572	773	17	34	401	201	4.970
AfIII	8648	11778	23.443	11558	220	629	6956	1862	46.044
AgeI	1130	2084	4.148	2031	53	112	1183	533	13.180
AlwNI	2491	4204	8.368	4163	41	131	2782	1074	26.558
Apal	406	738	1.469	719	19	49	390	194	4.797
ApaI	1663	2980	5.931	2907	73	125	1796	730	18.051
ApeKI	13902	14194	28.252	13910	284	782	8957	2190	54.154
ApoI	111478	7102	14.136	6871	231	549	4199	1390	34.372

Figure 2. Overview report of an Original RAD protocol simulation of the *C. elegans* reference genome.

Table 1. Description of column headers in the overview page of RApYDs simulation report.

Column headers	Description
Restriction enzyme	Restriction enzymes tested
Fragments after digestion	Number of fragments produced after enzymatic digestion. For ddRAD this accounts for the first and second digestion.
Fragments after size selection (RAD loci)	Number of fragments produced after size selection. This is equivalent to the number of RAD loci produced
Percent breadth of coverage	Percent of the input sequence length covered by the RAD loci
Single-copy RAD loci	Number of RAD loci found only once in the reference sequence. If at least one of the read pairs representing the RAD loci maps uniquely to the reference genome, the loci is considered to be single-copy.
Repetitive RAD loci	Number of RAD loci found multiple times in the reference sequence. If none of the read pairs representing the RAD loci maps uniquely to the reference genome, the loci is considered to be repetitive.
Repeat regions with RAD loci	Number of repeat regions in the reference sequence harboring RAD loci
RAD loci in specified feature	Number of RAD loci that are found within the specified sequence feature
Specified feature with RAD loci	Number of specified sequence feature in the reference harboring RAD loci
Percent specified feature with RAD loci	Percentage of specified sequence feature with RAD loci based on total number of specified feature

Table 2. Average running time of RApYDS (with histogram graphing skipped) on each species using a server with 16 CPU Cores of Intel(R) Xeon(R) CPU E5-2640 v2 @ 2.00GHz and 50GB of memory allocation.

Species	FASTA file size (MB)	Average Processing Time (sec)		
		With Alignment and Reference Indexing	With Alignment Only	Without Alignment
<i>Escherichia coli</i> ^a	4.50	16.89	14.56	6.41
<i>Caenorhabditis elegans</i> ^a	97.00	656.18	530.86	132.85
<i>Brassica napus</i> ^b	943.00	2093.52	722.63	356.56
<i>Aedes aegypti</i> ^b	1200.00	4092.68	1684.06	584.77
<i>Zea mays</i> ^b	2000.00	8291.83	4311.70	899.99
<i>Callithrix jacchus</i> ^b	2700.00	8063.47	2165.91	1288.30

^a RApYDS default parameters (original RAD, entire enzyme database)^b RApYDS ddRAD parameters (enzymes used: NlaIII-MluCI)

visualization, the user can specify the sizes of the DNA ladder rungs to display (Lane 1), and can also choose up to five enzymes for which the simulated banding patterns will be shown (Lanes 2 to 6). This particular view can be helpful to researchers who are more adept at assessing the utility of restriction enzymes based on gel electrophoresis profiles.

To give the users information on the distribution and density of the enzyme cut sites across the entire length of the reference sequence, the tool also outputs a view of the cut site distribution for each of the enzymes included in the simulation (Figure 4). This view can be magnified to focus on specific regions with greater resolution. A histogram

Figure 3. Gel visualization of the simulation results on *C. elegans* reference genome using an Original RAD protocol with BamHI, BstYI, PstI, SbfI and NotI as REs.

of the enzyme cut site density per binned region is also generated.

Sample application

In general, the users can do RAD simulations for a given target fragment size range to be generated from one or more REs available in the tool's database. Alternatively, if a particular RE is not included in the database, information on this RE can be manually added and subsequently used. It is also recommended for the users to have a prior idea of an appropriate reference genome to be used in estimating the distribution of target features across the entire genome length. For instance, in capturing polymorphic sites, the number of RAD loci may drastically change when using common cutter enzymes exclusively or in combination with rare cutter enzymes.

Certain resequencing studies also involve calling variants using whole genome data sets sequenced at low coverage depth in an attempt to reduce the overall sequencing cost. However, a RADseq approach can result in a similar cost reduction (O'Rourke *et al.*, 2011) without sacrificing coverage depth and variant loci density if an appropriate set of restriction enzymes is used. Specifically, enzymes that can generate a relatively even fragment length distribution across the entire genome are generally desirable (Roszik *et al.*, 2017). Thus, RapyDS

can be utilized to evaluate which restriction enzyme or restriction enzyme pair can generate the desired RAD loci density and distribution given a range of insert sizes and within certain genomic features (e.g., genes, exons) if an annotation file is supplied. By default, the tool will also output the number of single- and multi-copy RAD loci based on mapping information, which can also be a consideration in deciding which particular restriction enzyme to use.

To demonstrate the use of RapyDS, an original RAD simulation was implemented on the *C. elegans* reference genome using all the REs present in the tool's database. Figure 2 shows a portion of the results arranged alphabetically based on RE names. We can infer from these results that using *Acil* on *C. elegans* chromosome I produced the highest number of RAD loci (16,082), covering 32% of the reference sequence length with the default size selection range of 200-300 bp. Based on alignment results, *Acil* enzyme is predicted to yield mostly single-copy RAD loci (15,815), with only 1.66% (267) repetitive RAD loci detected and 5.32% (856) occurring at repeat regions. The number of sequenced repetitive RAD loci and/or RAD loci repeat regions can pose issues in downstream analysis that are heavily dependent on mapping and alignment approaches.

Figure 4. Cut site and RAD loci density distribution of the restriction enzyme Acc65I on chromosome I of *C. elegans*.

Future developments

There are certain biological factors not accounted for in the current implementation of RApYDS that may affect actual RADSeq experiments. For instance, certain restriction enzymes can be sensitive to the methylation state of a particular nucleotide (Mülhardt & Beese, 2007), which is usually not reflected in the reference sequence. Single nucleotide polymorphisms in the enzyme recognition sites can also result in allele drop-outs that can further lead to a systematic underestimation of the overall diversity in population studies, particularly for organisms with relatively high mutation rates (Andrews *et al.*, 2016). We will try to further incorporate some of these biological considerations in future releases of RApYDS in order to obtain simulation metrics that can more closely reflect the results from real experimental data sets.

Data availability

Underlying data

NCBI Assembly: AaegL5.0. RefSeq assembly accession: GCF_002204515.2.

NCBI Assembly: ASM584v2. RefSeq assembly accession: GCF_000005845.2

NCBI Assembly: B73 RefGen_v4. RefSeq assembly accession: GCF_000005005.2.

NCBI Assembly: Bra_napus_v2.0. RefSeq assembly accession: GCF_000686985.2.

NCBI Assembly: Callithrix jacchus-3.2. RefSeq assembly accession: GCF_000004665.1.

NCBI Assembly: WBcel235. RefSeq assembly accession: GCF_000002985.6.

Software availability

Source code available from: <https://github.com/pgcbioinfo/rapyds>.

License: GNU General Public License v3 (GPLv3).

References

- Andrews KR, Good JM, Miller MR, *et al.*: **Harnessing the power of RADseq for ecological and evolutionary genomics.** *Nat Rev Genet.* 2016; **17**(2): 81–92.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Baird NA, Etter PD, Atwood TS, *et al.*: **Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers.** *PLoS One.* 2008; **3**(10): e3376.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Davey JW, Blaxter ML: **RADSeq: next-generation population genetics.** *Brief Funct Genomics.* 2010; **9**(5–6): 416–423.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Davey JW, Cezard T, Fuentes-Utrilla P, *et al.*: **Special features of RAD Sequencing data: implications for genotyping.** *Mol Ecol.* 2013; **22**(11): 3151–3164.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Etter PD, Bassham S, Hohenlohe PA, *et al.*: **SNP discovery and genotyping for evolutionary genetics using RAD sequencing.** *Methods Mol Biol.* 2011; **772**: 157–178.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Herrera S, Reyes-Herrera PH, Shank TM: **Predicting RAD-seq Marker Numbers across the Eukaryotic Tree of Life.** *Genome Biol Evol.* 2015; **7**(12): 3207–3225.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lepais O, Weir JT: **SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches.** *Mol Ecol Resour.* 2014; **14**(6): 1314–1321.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics.* 2009; **25**(14): 1754–1760.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mülhardt C, Beese EW: **The Tools.** *Molecular Biology and Genomics.* Oxford: Academic Press. 2007; 37–63.
[Publisher Full Text](#)
- Meger J, Ulaszewski B, Vendramin GG, *et al.*: **Using reduced representation libraries sequencing methods to identify cpDNA polymorphisms in European beech (*Fagus sylvatica* L).** *Tree Genet Genomes.* 2019; **15**(1): 7.
[Publisher Full Text](#)
- O'Rourke SM, Yochem J, Connolly AA, *et al.*: **Rapid Mapping and Identification of Mutations in *Caenorhabditis elegans* by Restriction Site-Associated DNA Mapping and Genomic Interval Pull-Down Sequencing.** *Genetics.* 2011; **189**(3): 767–778.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peterson BK, Weber JN, Kay EH, *et al.*: **Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species.** *PLoS One.* 2012; **7**(5): e37135.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roszik J, Fenyőfalvi G, Halász L, *et al.*: ***In Silico* Restriction Enzyme Digests to Minimize Mapping Bias in Genomic Sequencing.** *Mol Ther Methods Clin Dev.* 2017; **6**: 66–67.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shendure J, Ji H: **Next-generation DNA sequencing.** *Nat Biotechnol.* 2008; **26**(10): 1135–1145.
[PubMed Abstract](#) | [Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research