A New Algorithm for Identifying Genome Rearrangements in the Mammalian Evolution

Juan Wang¹, Bo Cui¹, Yulan Zhao¹ and Maozu Guo²,³*

¹School of Computer Science, Inner Mongolia University, Hohhot, China, ²School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China, ³Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing, China

Genome rearrangements are the evolutionary events on level of genomes. It is a global view on evolution research of species to analyze the genome rearrangements. We introduce a new method called RGRPT (recovering the genome rearrangements based on phylogenetic tree) used to identify the genome rearrangements. We test the RGRPT using simulated data. The results of experiments show that RGRPT have high sensitivity and specificity compared with other tools when to predict rearrangement events. We use RGRPT to predict the rearrangement events of six mammalian genomes (human, chimpanzee, rhesus macaque, mouse, rat, and dog). RGRPT has recognized a total of 1,157 rearrangement events for them at 10 kb resolution, including 858 reversals, 16 translocations, 249 transpositions, and 34 fusions/fissions. And RGRPT has recognized 475 rearrangement events for them at 50 kb resolution, including 332 reversals, 13 translocations, 94 transpositions, and 36 fusions/fissions. The code source of RGRPT is available from https://github.com/wangjuanimu/data-of-genome-rearrangement.

Keywords: genome rearrangements, mammal, phylogenetic tree, evolution, algorithm

INTRODUCTION

The rapid development of sequencing technologies makes the phylogenetic analysis from the level of whole genome possible. A studied genome is represented as a line of conserved segments (called syntenic blocks). The genome rearrangements of species are changes of syntenic block orderings and losing of sequence blocks. These events include reversal, translocation, transposition, fusion, fission, and so on (Xu et al., 2017; Cheng et al., 2019; Dong et al., 2018). The research on genome rearrangements is mainly three aspects.

One is the computation of evolutionary distance between two species by considering genome rearrangements. Researchers have proposed a lot of metric for measuring the dissimilarity of evolution between species and a large amount of algorithms for computing the metrics. The breakpoint distance is the minimum rearrangement operations transforming one genome to the other genome, which is computed by means of breakpoint graph (Blanchette et al., 1997; Sankoff and Blanchette, 1998). There are lots of algorithms for computing breakpoint distance. In 1995, Hannenhalli and Pevzner put forward an algorithm with O(n³) time complexity to compute the breakpoint distance just considering reversal events (Hannenhalli and Pevzner, 1999). Later, Kaplan improved the algorithm to time complexity O(n²) (Kaplan et al., 2000). In 1996, Hannenhalli designed an algorithm with O(n³) time complexity to compute it by
considering translocation events (Hannenhalli, 1995). In 2001, Zhu et al. improved the algorithm to time complexity \(O(n^2 \log n)\) (Zhu and Ma, 2002). And then Zhu et al. devised an algorithm with \(O(n^2)\) time complexity (Liu et al., 2004). The DCJ distance is introduced by Yancopoulos et al. (Sophia et al., 2005), which uses the double cut and join (DCJ for short) operation to model rearrangement events, such as reversal, translocation, transposition, fusion, and fission in an unified way. Yancopoulos et al. first propose a method to compute the DCJ distance by considering only translocations and reversals on linear chromosomes (Sophia et al., 2005). Paper (Liu et al., 2006) has proposed an \(O(n^2)\) time algorithm to compute the distance by considering the fusions and fissions between circular unsigned chromosomes. Unimog (Hilker et al., 2012) is software for computing DCJ distance which implements lots of algorithms (Erdös et al., 2011; Jakub et al., 2011). SoRT is a tool to compute breakpoint distance and the DCJ distance for linear/circular multi-chromosomal gene orders (Yen-Lin et al., 2010). SCJ distance (Feijão and Meidanis, 2011) is defined using the single cut and join (SCJ for short) operations, which is in analogy to DCJ measure. The distance can be computed by a speedily computable.

Two is the reconstruction of the ancestral gene orders by using the genomes of extant species. Ma et al. (Ma et al., 2006) use maximum parsimony principle to recover reliably ancestral genomes starting from phylogenetic tree and adjacent genes in genome and make the probabilistic reconstruction accuracy analysis for the six mammalian genome (human, mouse, rat, dog, opossum, and chicken) based on the improved Jukes–Cantor model. PMAG utilized the Bayesian theorem in the probabilistic framework to infer ancestral genomes (Yang et al., 2014). Multiple Genome Rearrangements (MGR) recovers the ancestral genome by minimizing the rearrangement distance (Bourque and Pevzner, 2002). Multiple Genome Rearrangements and Ancestors (MGRA) is developed to reconstruct ancestral genomes based on multiple breakpoint graphs and is used to analyze rearrangement evolutionary events of seven mammalian genomes (human, chimpanzee, macaque, mouse, rat, dog, and opossum) (Alexseyev and Pevzner, 2009). Decostar (Duchemin et al., 2017) is a software which reconstructs neighborhood relations of ancestral genes aiming at reconstructing the organization of ancestral genomes.

Three is the recognition of the rearrangement events of existing species. Efficient Method to Recover Ancestral Events (EMRAE) is an algorithm which can recognize rearrangement events in evolution described by phylogenetic tree by means of adjacent genes in genomes (Zhao and Bourque, 2009).

**MATERIALS AND METHODS**

**Preliminaries**

A genome is composed of several chromosomes, and each chromosome is an ordering of syntenic blocks. For convenience, each syntenic block is recorded by an integer, so a chromosome is represented by a signed permutation \(X = c_1c_2 \cdots c_n\), where \(c_i(1 \leq i \leq n)\) is an integer representing a syntenic block, its sign is assigned with the orientation that is either positive (recorded by \(c_i\)) or negative (recorded by \(-c_i\)). The chromosome \(X = c_1c_2 \cdots c_n\) is the same as \(-X = c_n \cdots c_2c_1\).

A reversal \(r(i, j) (i \leq j)\) converts chromosome \(X = c_1c_2 \cdots c_i \cdots c_j \cdots c_n\) into a new chromosome \(X' = c_1c_2 \cdots c_{i-1} \cdots c_{j+1} \cdots c_n\), where the reversal is from \(c_i\) to \(c_j\).

A translocation event breaks two chromosomes into four segments and then reconnects them into two new chromosomes. Given two chromosomes \(X = X_1X_2\) and \(Y = Y_1Y_2\), where \(X_1x_2x_3 \cdots x_nX_2 = x_1x_4 \cdots x_{n-1}x_n\) and \(Y_1y_2y_3 \cdots y_n = y_1y_4 \cdots y_{n-1}y_n\), a translocation is represented by \(i/l(j)\). \(X_1\) and \(Y_1\) are exchanged to form two new chromosomes \(X' = Y_1X_1X_2\) and \(Y' = X_1Y_2\), and \(X_2\) and \(Y_2\) are exchanged to form two new chromosomes \(X'' = X_2Y_1\) and \(Y'' = X_1Y_2\).

A transposition event is to exchange two adjacent fragments on one chromosome into a new chromosome. A transposition is represented by \(tp(i, j)\), i.e., the fragment \(c_i \cdots c_j\) of one chromosome inserted into after \(c_i\). If \(c_i\) is on the same chromosome \((k > j)\), then the transposition \(tp(i, j)\) is called intra-chromosomal; otherwise, it is inter-chromosomal. Given a chromosome \(X = c_1c_2 \cdots c_{i-1}c_i \cdots c_{j-1}c_j \cdots c_{n-1}c_n\) and an intra-chromosomal transposition, \(X\) is converted into \(X' = c_1c_2 \cdots c_{i-1}c_{i+1} \cdots c_{j-1}c_j \cdots c_{n-1}c_n\).

A fusion event is to connect two chromosomes into a new chromosome. The fusion acting on chromosomes \(X_1\) and \(X_2\) is represented by \(u(X_1, X_2)\) and forming a new chromosome \(X_1' = X_1X_2\) or \(X_2' = X_1X_2\). A fusion is to split a chromosome into two new chromosomes. A fission acting on chromosomes \(X_1\) and \(X_2\) is represented by \(f(X)\) and forming two new chromosomes \(X_1' = X_1X_2\) and \(X_2' = X_1X_2\) (where \(X_1\) and \(X_2\) are non-empty segments).

An adjacency \(a(c_i, c_{i+1})\) of genome \(X\) is two adjacent integers in one chromosome of \(X\). \(a(c_i, c_{i+1})\) is the same as \((-c_i, c_{i+1})\). For example, all adjacencies on chromosome \(X = 1,2,3,4\) are \(a(1,2), a(2,3), a(3,4)\). For a set of genomes \(S\), an adjacency \(a\) is effective w.r.t. \(S\) if it belongs to at least one genome and not all genomes. For example, two uni-chromosomal genomes \(G_1\) and \(G_2\), the chromosome \(X = 1,2,3,4\) of \(G_1\) and the chromosome \(Y = 1,3,24\) of \(G_2\), then all effective adjacencies w.r.t. \(G_1\) and \(G_2\) are \(a(1,2), a(2,3), a(3,4), a(1,3), a(-2,4)\).

**EMRAE**

Given a phylogenetic tree \(T\) describing the evolution of the genomes \(G\), EMRAE first computes all effective adjacencies w.r.t. \(G\). Then, it predicts the rearrangement events for each edge of \(T\) by means of inference rules (will be introduced in the following).

**Figure 1** shows a reversal \(r(2,3)\) during the evolution from \(A\) to \(B\), where \(A\) and \(B\) are two uni-chromosomal genomes, and the chromosomes are \(X = 1,2,3,4\) and \(Y = 1,3,2,4\), respectively. The set of genomes will be divided into two subsets recorded by \(S_A\) and \(S_B\) after removing the edge \(e\) from \(T\). Suppose there is not any rearrangement events inside \(S_A\) and \(S_B\). Then, adjacencies \(a(1,2)\) and \(a(3,4)\) can be found in each genome of \(S_A\) and not in any one genome of \(S_B\); \(a(1,−3)\) and \(a(−2,4)\) can be
found in each genome of $S_b$ and not in any one genome of $S_a$.

In turn, we can utilize the four adjacencies $a(1, 2), a(3, 4), a(1, -3),$ and $a(-2, 4)$ to identify a reversal $(2, 3)$ occurring on the edge $e$. The EMRAE method infers the rearrangement events by means of the similar rules.

Let $e = (A, B)$ be an edge of $T$, $G=[G_1, G_2, \ldots, G_m]$ the genomes of leaves, and $a_1, a_2, \ldots, a_n$ the children of $A$ and $b_1, b_2, \ldots, b_n$ the children of $B$. EMRAE first selects a number of adjacencies as candidate adjacencies $Ca(e,A)$ for edge $e$ and node $A$ according the following steps.

1. Find the adjacencies are in each genome of $S_i$ and not in any one genome of $S_b$, then put them to $Ca(e, A)$;
2. If an internal node, find all edges connected with $A$ except $e$ and record them with $e_1, e_2, \ldots, e_k$. For each $e_i=(u_i, A), (1 \leq i \leq k)$, $G$ can be divided into two parts after removing $e_i$, $S_u$ is the part not including $A$.
   a. Find the adjacencies that are in one genome of each $S_u$ ($1 \leq i \leq k$) and not in any one genome of $S_b$, then put them to $Ca(e, A)$;
   b. Compute $Ca(e, u)$ and $Ca(e, u)(1 \leq i \leq k)$. For each one $Ca(e, u)$, find the adjacency $a_i$ from $Ca(e, u)$, such that $a_i$ is not overlap gene with any one adjacency in $Ca(e, u)$, $a_i$ has overlap gene with one adjacency $a_j$, in each $Ca(e, u)(1 \leq j \leq k)$, $a_i$ has overlap gene with at least one adjacency in $Ca(e, u)$, then put $a_i \in \delta do5(1)$ to $Ca(e, u)$.

EMRAE then infers rearrangement from $Ca(e, A)$ and $Ca(e, B)$ for edge $e = (A, B)$ with the help of inference rules in the following section. From the definitions of genome rearrangements, we find that each genome rearrangement can change several adjacencies. For example, each reversal $(i, j)(i \leq j)$ can change two adjacencies $a_i=a(c_{i-1}, c_i)$ and $a_j=a(c_{j+1}, c_j)$ into $b_i=a(c_{i-1}, -c_i)$ and $b_j=a(-c_{j+1}, c_j)$. Based on those facts, we obtain the inference rules introduced in the following section.

**Inference Rule**

Let $e = (A, B)$ be an edge of the phylogenetic tree $T$. Given adjacencies $a_1=a(c_{i-1}, c_i), a_2=a(c_{j+1}, c_j)$ in $Ca(e,A)$ and $b_1=a(c_{i-1}, -c_i), b_2=a(-c_{j+1}, c_j)$ in $Ca(e,B)$, EMRAE infers a reversal $r(i,j)$ from $A$ to $B$ if all genomes are uni-chromosomal or $a_1, a_2$ are in the same chromosome in $S_A$ and $b_1, b_2$ are in the same chromosome in $S_B$. Otherwise, we infer a translocation $tl(i,j)$. Similarly, given
RESULTS

All of the experiments were performed on a computer with Intel Vostro 14 2.0 GHz CPU, 4 GB RAM, and 500 GB Hard Disk Drives (HDD). The operating system was Win10 64 bit with Java 1.6 installed. RGRPT was written in Java.

We tested RGRPT with both simulated data and the practical data (i.e., real biological data) introduced by the following section.

Simulated Data

Here, we start with an uni-chromosomal genome as the ancestor, and it evolves along the phylogenetic tree with \( n \) taxa whose topology sees the Figure 3.

We generate two simulated data sets in order to test the affectivity of RGRPT. One of them is created from the phylogeny only with reversals events. The other data set is generated from the phylogeny with kinds of events, including reversals, translocation, transposition, fusion, and fission, and the quantity of those events is in a certain ratio. The two data sets can test the ability of methods to recover the simple and the complex evolution histories. First data set is created just using reversal events. Since the reversal on only one gene is rare (Korbel et al., 2007), we set the ratio of reversal on one gene and on more than one gene as 1:3. The number of leaves is from 3 to 10 with step 1. For each number of leaves, the ancestor genome with \( m \) gene, where \( m \) from 50 to 150 with step 10. Each edge will happen \( k \) reverse, where \( k \) is random integer number from 3 to 10. So, there are 11 groups data for each leaf number. Sensibility is the percentage of correctly predicted events in all practical events. Specificity is the percentage of correctly predicted events in all predicted events. We compute the sensibility and specificity for RGRPT and EMRAE for each group data. Table 1 shows the average sensitivity and specificity for each leaf number. The second column of the table records the number of all events, and its last row records the average values.

Table 1 shows that RGRPT achieves higher sensibility than EMRAE, and RGRPT achieves comparable specificity with EMRAE. Obviously, RGRPT can distinguish more actually occurred events than EMRAE. So, the experimental results show that the RGRPT is more efficient than EMRAE for predicting reversal events.

Second data set is generated by using all events, i.e., reversal, translocation, transposition, fusion, and fission. The reversals are generally more than the other rearrangement events. The fusions and the fissions are very rare; so, we record the number of the two events together. Here, we set the ratio of those events as 10:2:2:0.1. The ancestor genome has 5 chromosomes and each chromosome with 100 genes. The ancestor genome evolves along the topology with four leaves (see Figure 3). Each edge happen \( k \) events, where \( k \) is random number from 1 to \( \mu \) and \( \mu \) is 6, 12, 18, and 24. For each \( \mu \), it runs 10 times; so, we can obtain 10 groups data for each \( \mu \). Table 2 shows the average of 10 groups data for each \( \mu \). This table indicates that the RGRPT is more efficient than EMRAE for predicting all events.

Practical Data

The practical data is from the paper (Zhao and Bourque, 2009). It contains six mammalian genomes, i.e., human, chimpanzee, rhesus monkey, mouse, voles, and dog. The data are created from two different levels of resolution 10 kb and 50 kb. Figure 4 is the tree describing the phylogeny of species. The results are shown in Tables 3 and 4. EM and RG represent EMRAE and RGRPT respectively, and Rev, Tloc, Tran, Fus, and Fis represent reversal, translocation, transposition, fusion, and fission.

| Leaves | Reversal | Sensibility | Specificity |
|--------|----------|-------------|-------------|
|        |          | EMRAE       | RGRPT       | EMRAE       | RGRPT       |
| 3      | 24       | 64%         | 76%         | 89%         | 90%         |
| 4      | 39       | 66%         | 76%         | 94%         | 94%         |
| 5      | 45       | 61%         | 72%         | 92%         | 93%         |
| 6      | 59       | 57%         | 66%         | 90%         | 90%         |
| 7      | 69       | 54%         | 65%         | 92%         | 91%         |
| 8      | 79       | 59%         | 80%         | 92%         | 92%         |
| 9      | 92       | 55%         | 63%         | 90%         | 90%         |
| 10     | 104      | 55%         | 62%         | 89%         | 89%         |
| Mean   |          | 58.7%       | 70%         | 91%         | 91.1%       |
translocation, transposition, fusion, and fission, respectively. Each row in the table records the ancestor rearrangement events of the edge. For example, the values in the human row are the rearrangement events from D to human; the values in MR row are the rearrangement events from A and B.

At 10 kb resolution, the RGRPT algorithm predicts 1,157 ancestor rearrangement events, including 858 reversals, 16 translocations, 249 transpositions, and 34 fusions and fissions. It identifies 48 rearrangement events more than the EMRAE. The reversal events are in the majority in all predicted events. At 50 kb resolution, the RGRPT algorithm predicts 475 ancestor rearrangement events, including 332 reversals, 13 translocations, 94 transpositions, and 36 fusion and fissions. RGRPT identifies 21 rearrangement events more than EMRAE algorithm. The rearrangement events identified in the rat

### TABLE 2 | Results of EMRAE and recovering the genome rearrangements based on phylogenetic tree algorithms in predicting all events.

| Events of each edge | All events | Sensibility | Specificity |
|---------------------|------------|-------------|-------------|
|                     |            | EMRAE       | RGRPT       | EMRAE       | RGRPT       |
|                      |            | Sensibility | Specificity |
|                     |            | 75.8%       | 86.7%       | 95.8%       | 96.2%       |
|                     |            | 74.2%       | 80.3%       | 97%         | 96.5%       |
|                     |            | 53.5%       | 58.1%       | 95.4%       | 96.7%       |
|                     |            | 47.7%       | 50.5%       | 94.9%       | 94.1%       |
| Mean                |            | 62.8%       | 68.7%       | 95.8%       | 95.9%       |

### TABLE 3 | Genome rearrangement predictions of EMRAE and recovering the genome rearrangements based on phylogenetic tree at 10 kb resolution.

| Species | Rev | Tloc | Tran | Fus/Fis | Total events |
|---------|-----|------|------|---------|--------------|
|         | EM  | RG   | EM   | RG      |              |
| Human   | 12  | 13   | 0    | 0       | 16           |
| HC      | 29  | 32   | 0    | 0       | 44           |
| HCP     | 83  | 84   | 0    | 0       | 93           |
| Chimpanzee | 17  | 19    | 0    | 0       | 25           |
| Rhesus | 49  | 50   | 0    | 0       | 90           |
| Mouse  | 90  | 95   | 3    | 3       | 108          |
| Rat     | 227 | 233  | 0    | 0       | 357          |
| MR      | 140 | 143  | 2    | 3       | 151          |
| Dog     | 184 | 189  | 10   | 10      | 225          |
| Total   | 831 | 858  | 15   | 16      | 1,109        |

### TABLE 4 | Genome rearrangement predictions of EMRAE and recovering the genome rearrangements based on phylogenetic tree at 50 kb resolution.

| Species | Rev | Tloc | Tran | Fus/Fis | Total events |
|---------|-----|------|------|---------|--------------|
|         | EM  | RG   | EM   | RG      |              |
| Human   | 2   | 2    | 0    | 0       | 3            |
| HC      | 19  | 19   | 0    | 0       | 24           |
| HCP     | 27  | 29   | 0    | 0       | 34           |
| Chimpanzee | 17  | 19    | 0    | 0       | 25           |
| Rhesus | 22  | 23   | 0    | 0       | 29           |
| Mouse  | 25  | 27   | 3    | 3       | 33           |
| Rat     | 128 | 131  | 0    | 0       | 198          |
| MR      | 41  | 42   | 2    | 2       | 45           |
| Dog     | 48  | 47   | 7    | 8       | 74           |
| Total   | 322 | 332  | 12   | 13      | 454          |
edge are mostly in all edges either at 10 kb resolution or at 50 kb resolution. The syntenic blocks of genomes at 10 kb resolution are more than the syntenic blocks of genomes at 50 kb resolution. The fact reduces the recognized rearrangement events at 10 kb resolution that are more than the recognized rearrangement events at 50 kb resolution. Experiments show that RGRPT can recover more ancestor events than EMRAE.

**DISCUSSION**

This paper proposes a new method, RGRPT, to infer ancestor rearrangement events. RGRPT takes a phylogenetic tree describing the evolution of species and the genomes of species as input. Experiments on the simulated data and practical data show that RGRPT is more efficient than EMRAE and can recover more ancestor rearrangement events than EMRAE. RGRPT provides a method for us to research the genome rearrangement of species. We can use RGRPT to recognize the ancestral genome rearrangement for the evolution of other species in future (Tian et al., 2018).

**REFERENCES**

Alekseyev, M. A., and Pevzner, P. A. (2009). Breakpoint graphs and ancestral genome reconstructions. *Genome Res.* 19 (5), 943–957.

Blanchette, M., Bourque, G., and Sankoff, D. (1997). Breakpoint phylologies. *Genome Inform. Ser. Workshop Genome Inform.* 8, 25–34.

Bourque, G., and Pevzner, P. A. (2002). Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* 11 (1), 26–36.

Cheng, L., Yang, H., Zhao, H., Pei, X., Shi, H., Sun, J., et al. (2019). Metsigdis: a manually curated resource for the metabolic signatures of diseases. *Briefings Bioinf.* doi: 10.1093/bib/bbx103.

Dong, S., Zhao, C., Fei, C., Liu, Y., Zhang, S., Hong, W., et al. (2018). The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *Bmc Genomics* 19 (1), 614–626.

Duchemin, W., Anselmetti, Y., Patterson, M., Ponty, Y., Brard, S., Chauve, C., et al. (2017). Decostar: Reconstructing the ancestral organization of genes or genomes using reconciled phylogenies. *Genome Biol. Evol.* 9 (5), 1312–1319.

Erdős, P. L., Soukup, L., and Stoye, J. (2011). Balanced vertices in trees and a simpler algorithm to compute the genomic distance. *Appl. Math. Lett.* 24 (1), 82–86.

Feijão, P., and Meidanis, J. (2011). Scja breakpoint-like distance that simplifies several rearrangement problems. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8 (5), 1318–1329.

Hannenhalli, S. (1995). Polynomial-time algorithm for computing translocation distance between genomes. *Discrete Appl. Math.* 71 (1–3), 137–151.

Hannenhalli, S., and Pevzner, P. A. (1999). Transforming cabbage into turnip:polynomial algorithm for sorting signed permutations by reversals. *J. Acm* 46 (1), 1–27.

Hilker, R., Sickinger, C., Pedersen, C. N., and Stoye, J. (2012). Unimog—a unifying framework for genomic distance calculation and sorting based on dcj. *Bioinformatics* 28 (19), 2509.

Jalalvand, K., Robert, W., Braga, M. D. V., and Jens, S. (2011). Restricted dcj model: rearrangement problems with chromosome reincorporation. *J. Comput. Biol.* 18 (9), 1231–1241.

Kaplan, H., Shamir, R., and Tarjan, R. E. (2000). Faster and simpler algorithm for sorting signed permutations by reversals. *SIAM J. Comput.* 29 (3), 880–892.

Korbel, J. O., Urban A. E., Affourtit J. P., Godwin B., Grubert F., Simons J. F., et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318 (5849), 420–426.

**DATA AVAILABILITY STATEMENT**

Publicly available datasets were analyzed in this study. These data can be found here: https://github.com/wangjuanimu/data-of-genome-rearrangement.

**AUTHOR CONTRIBUTIONS**

JW proposed and implemented the RGRPT method. JW and BC designed all experiments. All authors participated in the designing the algorithm and writing the paper.

**FUNDING**

The work was supported by the National Natural Science Foundation of China (61661040, 61661039, 61571163, 61532014, 61671189, 91735306, 61751104); the National Key Research and Development Plan Task of China (Grant No. 2016YFC0901902).