Mobile Networks and Applications
https://doi.org/10.1007/s11036-020-01726-w

Design and Research of Intelligent Question-Answering(Q&A) System Based on High School Course Knowledge Graph

Zhijun Yang[1] · Yang Wang[2] · Jianhou Gan[3] · Hang Li[4] · Ning Lei[2]

Accepted: 6 December 2020
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract
Question answering is an indispensable link in high school teaching. Through question answering, on the one hand, it can solve students' learning doubts, on the other hand, it can provide teachers with teaching feedback. However, through the investigation and research, it is found that with the expansion of student size, the effect of question answering in high school is not satisfactory. This paper analyzes the current situation of question answering in high school, and designs an intelligent question answering system for high school teaching based on constructivism learning theory and cognitive structure learning theory. The system, which is the first innovative application in the field of high school teaching, integrates knowledge graph technology and intelligent question answering technology, introduces big data technology. It can solve students' questions in time and accurately, link the knowledge points related to the questions to help students construct knowledge network graph, and the big data technology is used to analyze the students' questioning behavior and to predict students' learning behavior in order to feedback the teaching effect.

Keywords High school · Knowledge graph · Intelligent question answering · Big data

1 Introduction

Education informatization is an important field and breakthrough of education industry in China. The country, schools and enterprises pay great attention to it. In the practice of "Internet + teaching", all kinds of schools at all levels actively use information technology to make teaching methods scientific, informative, intelligent and multimedia. At present, the rapid development of new generation of information technology, represented by artificial intelligence, cloud computing and big data, is pushing education informatization to a new height, which will bring new changes to education and learning. On April 13, 2018, *the action plan of educational informatization 2.0* issued by the Ministry of education of the people's Republic of China was formally put forward. In the plan, it is emphasized to promote the application of information technology in teaching, management and other aspects on the basis of artificial intelligence, big data, Internet of things and other emerging technologies [1, 2], relying on various information equipment and networks. Under this policy background, this paper designs an intelligent question answering system based on the knowledge graph of high school course, which is oriented to high school teaching and promotes the deep integration of information technology and teaching.

2 The status of question answering in high school

The basic links of the complete teaching work include preparing lessons, attending classes, arranging homework, question answering, teaching evaluation and so on. The main task of

✉ Jianhou Gan
ganjh@ynnu.edu.cn

Zhijun Yang
353738698@qq.com

Yang Wang
1358661928@qq.com

Hang Li
hang.li4@uq.net.au

Ning Lei
839259966@qq.com

[1] The Academy for Educational Science Research, Educational Department of Yunnan Province, Yunnan University, Kunming, China

[2] College of Vocational and Continuing Education, Yunnan University, Kunming, China

[3] Key Laboratory of Education Informalization for Nationalities of Ministry of Education, Yunnan Normal University, Kunming, China

[4] The University of Queensland, Brisbane, Australia

question answering is to answer the incomprehension of students in their study [3]. It is a way to fill in the gaps of classroom teaching, an active examination of teaching effect, and an important way to improve teaching. According to the investigation, there are three main forms of question answering in high school: 1 Choosing evening self-study, recess and other free time to ask teachers questions face to face; 2 Using Search Engines to get answers. 3 Using QQ, E-mail and other communication tools to communicate with teachers online. However, the analysis shows that there are some outstanding problems of question answering in high school:

2.1 The accuracy of answers is poor

With the development of the Internet, most students choose to use search engines to get answers. This method is not limited by time and space, and can get the answer in a short time. However, the disadvantage of this method is that there may be a lot of search results, which need to be filtered to get the results they need. In addition, high school students are weak in information recognition, so the accuracy of the answers is poor.

2.2 The timeliness of obtaining answers is weak

Some high school students have strong self-esteem and are embarrassed or afraid to ask questions to teachers in person. They will choose to use communication tools such as QQ, WeChat to communicate with teachers online. However, this method requires both parties to be online at the same time, otherwise students' problems cannot be solved in time and students' learning enthusiasm will be inhibited.

2.3 The efficiency of question answering is low

Through the investigation, it is found that many high schools set evening self-study as a special tutoring and answering time, students can ask teachers when they have questions during this time. However, the knowledge points in high school are relatively fixed, and students' problems are mostly similar. In this case, teachers may be asked the same question many times, which increases teachers' workload and reduces their work efficiency.

2.4 The comprehensiveness of answers is not enough

Knowledge points in high school are relatively fragmented, but there is a strong correlation between knowledge and knowledge. However, at present, no matter which way of question answering, only isolated knowledge points can be shown to students. In this way, students can not make deep and comprehensive connections between knowledge points, and the common problems may not be solved next time.

2.5 Lack of feedback mechanism for question answering

Students are the main body of teaching. By question answering, teachers can learn about students' learning, adjust teaching plans and improve teaching quality. However, there is no systematic feedback mechanism in the q&a work in high schools, and teachers have not systematically collected and analyzed the data of students' questions.

In view of the problems existing in high school question-answering activities and the importance of question answering in high school learning, this paper designs an intelligent q&a system based on high school course knowledge graph to help students timely and accurately answer their doubts and establish a knowledge network, to help teachers better check the teaching effect and improve the teaching.

3 Design and implementation of system

In recent years, the combination of knowledge graph and intelligent q&a is a hot topic, but the combination and application of the two are rare in the field of education [4, 5]. Innovation point of this article is, taking the high school question-answering activities as the starting point, on the basis of intelligent q&a system, using the knowledge graph technology to feedback structured knowledge points and related questions to learners in the form of knowledge trees [6, 7], so as to facilitate knowledge modeling, accurate positioning and in-depth acquisition of knowledge. In addition, this paper integrates big data technology and uses data mining algorithm to analyze students' question-answering behaviors, so as to better provide teachers with teaching feedback.

3.1 The work flow of the system

This system can be divided into three layers: user layer, analysis layer and data layer, which mainly includes four modules: problem reception, problem processing and matching, answer generation [8, 9] and question data analysis. The work flow of the system is shown in Fig. 1. First of all, students input problem in a natural language, then the system divides the questions into words, identifies named entities, and matches the identified entity information with the knowledge points in the knowledge graph database, so as to retrieve the answer with the highest similarity and the relevant knowledge structure diagram. Finally, the system uses big data technology to collect and analyze students' questioning behaviors, and return students' learning mastery to teachers.

3.2 Design and construction of knowledge graph

Knowledge graph is a structured semantic knowledge base, which is used to describe concepts and their relationships in

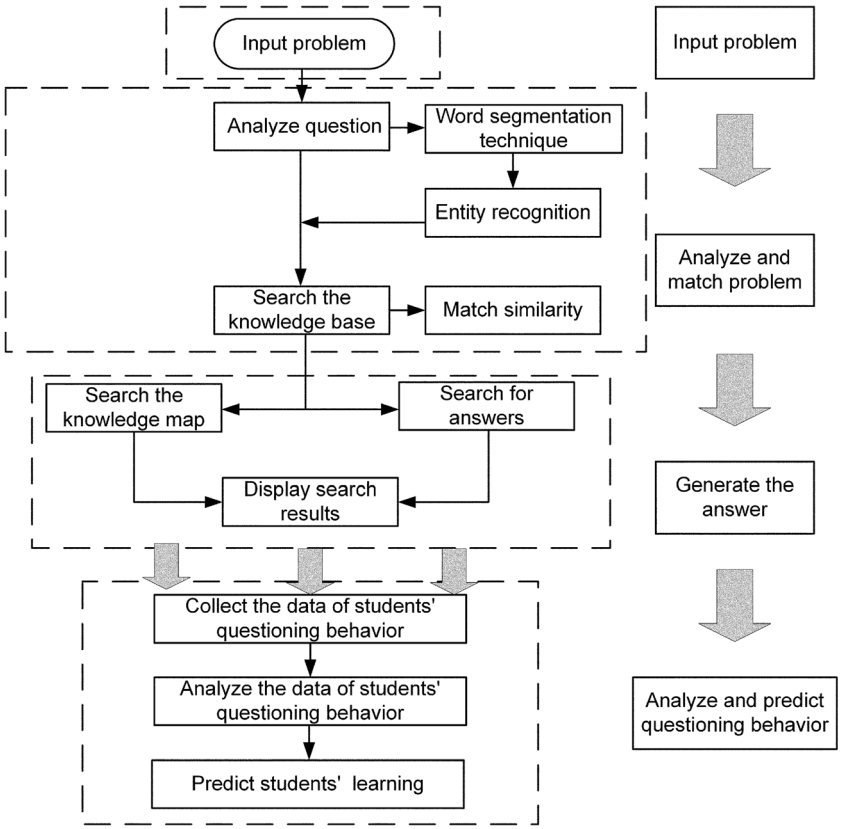

Fig. 1 Flow chart of system

the physical world in the form of graphs [10–12]. Its basic units are "entity-relation-entity" and "entity-attribute-value" [11]. According to the logical structure of knowledge graph, the ontology structure and construction process of knowledge graph must be clarified when constructing knowledge graph.

3.2.1 Ontology design

Because it is oriented to high school teaching, this system will construct high school course knowledge graph. To define the ontology structure of the course knowledge graph is to define the minimum concepts that the pattern layer in the course knowledge graph should contain and the relationship between them [13]. In this study, subject knowledge points are regarded as the knowledge units at the lowest level of the course knowledge graph, and the knowledge units of a course are represented as three levels: chapter, section and knowledge point. Furthermore, three kinds of relations among knowledge units are defined: inclusion relation, order relation and correlation relation. Figure 2 is the ontology structure of the high school course knowledge graph designed in this paper.

3.2.2 The construction of knowledge graph base

Based on the above ontology structure, the knowledge graph base of each subject is constructed. This process is divided into three steps: knowledge acquisition, knowledge representation and knowledge storage [12]. Firstly, this study gets knowledge from subject materials and syllabuses, subject area experts, teachers and students, and then, named entities and extraction relations are identified, and triples are used to represent knowledge and its relations. Next, knowledge integration is carried out according to the three relations defined in the ontology structure. Finally, the graph database Neo4j is used to store the knowledge graph and form the knowledge graph base.

3.3 Intelligent Q&a Technology based on knowledge graph

Intelligent q&a technology is a computer technology that can answer the questions raised by users in natural language as concisely and accurately as possible through the network. Intelligent q&a technology based on knowledge graph mainly includes four core modules: word segmentation, named entity recognition, problem similarity matching and query based on knowledge graph [14]. In this system, Reverse Maximum Matching(RMM) method is used for word segmentation, Conditional Random Fields(CRF) is used to identify named entities, TFIDF is used to calculate the similarity between extracted entities and knowledge points of the knowledge

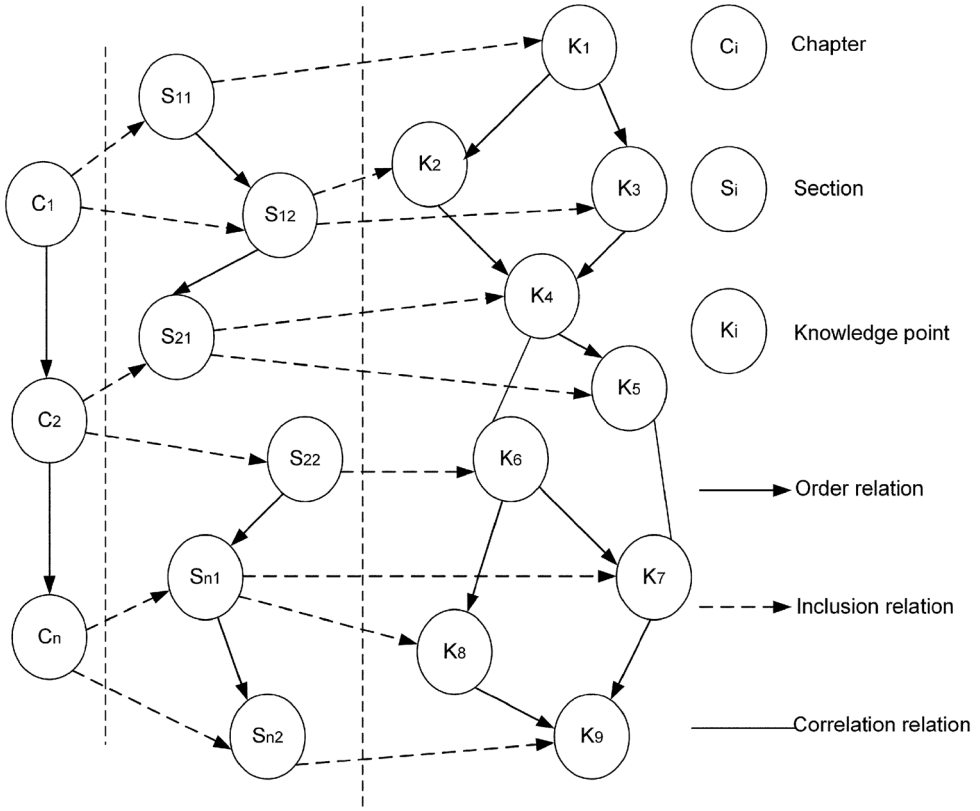

Fig. 2 Ontology structure of high school course knowledge graph

graph, and finally, Cypher statement provided by Neo4j is used to import data and query graph data.

3.3.1 Reverse maximum matching method

RMM is a word segmentation algorithm based on string matching. In RMM, for the problem statement, the pointer scans the string in reverse loop from right to left, and extracts the keyword table in the background for comparison. If the word exists in the keyword table, the matching word is extracted as the keyword for segmentation.

3.3.2 Conditional random fields

In this study, high school course knowledge is divided into knowledge units according to chapters, sections and knowledge points, which are the entities of course knowledge graph. This system uses the machine learning method based on statistical learning, namely CRF method, to identify the knowledge point entities in the problem.

3.3.3 TFIDF method based on vector space model

The identified entity information is matched with the knowledge points of the knowledge graph base to retrieve the answer with the highest similarity. The system uses TFIDF algorithm to calculate the similarity. First, Term frequency(TF) and inverse document frequency(IDF) multiplication to gain the weight of the entity vector [15], then weighted sum of all entity vectors to obtain the vector of the problem, and next calculating the Cosine and Euclidean distance between knowledge point vector and problem vector, finally, taking the average as the similarity, and returning the knowledge point with the highest similarity and related knowledge network diagram as answers.

3.3.4 Query based on knowledge graph

This system uses Cypher, the query language of Neo4j graphics database, to query knowledge graph base [16]. Firstly, generating the Cypher statement according to the entity name, entity category, and relationship name, then querying the knowledge graph database according to the SQL syntax, and finally returning the question answer and related knowledge network diagram.

3.4 The application of big data technology

In September 2015, the State Council issued *the action plan for promoting the development of big data*, in which the

requirements for the development of education big data were clearly put forward, and it was clearly emphasized to give full play to the important role of big data in education and teaching [17]. Under the background of this policy, in order to better play the role of teaching feedback of question answering, this study constructed a big data processing and analysis model based on students' questioning behavior, as shown in Fig. 3.

Firstly, the system collects students' questioning data in real time, and then analyzes the data by means of variance analysis, correlation analysis, factor analysis, cluster analysis and principal component analysis [18]. Secondly, prediction models and simulation technology are used to predict students' learning behaviors. Finally, based on the big data visualization technology, the visualization presentation and interaction of the analysis results are realized.

4 Teaching application mode

The design and development of any educational system can not be separated from the guidance of educational theory. In addition, we need to build a teaching application scene under the guidance of scientific and perfect education theory.

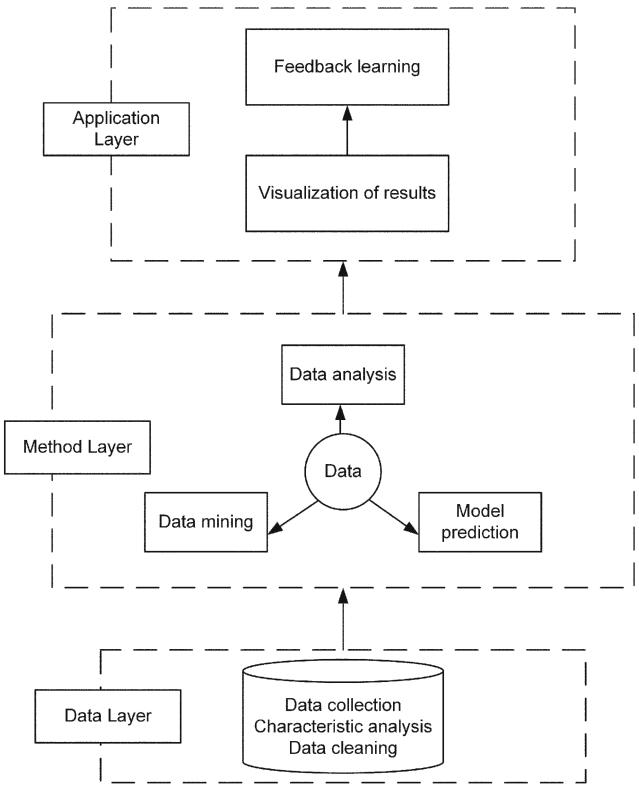

Fig. 3 Big data processing and analysis model based on students' questioning behavior

4.1 Theoretical basis

This study takes learners as the center, starts from learners' learning needs and cognitive characteristics, and takes constructivism learning theory and cognitive structure learning theory as the theoretical support of this system.

4.1.1 Constructivism learning theory

Constructivism learning theory takes learners as the center, and holds that students are the main body of cognition and the active constructors of knowledge meaning, while teachers are only the promoters of students' meaning construction. The theory emphasizes creating learning situations for students. In this situation, students can construct knowledge independently, understand the internal relationship of knowledge deeply, and form a solid knowledge schema in the brain.

According to Piaget's theory of cognitive development stage, senior high school students have been in the stage of formal operation, in which they have been able to independently monitor and introspect their thinking activities. Starting from the cognitive characteristics of senior high school students and according to the learning theory of constructivism, the system strives to create a problem situation for students that is conducive to independent learning. In this situation, it takes "problem" as the starting point to stimulate students' desire for exploration and enthusiasm for learning.

4.1.2 Cognitive structure learning theory

Bruner puts forward the theory of cognitive structure learning, which points out that students are not passive knowledge receivers, but active information processors. Bruner attaches great importance to students' mastery of subject knowledge structure. He believes that students' mastery of the knowledge structure of the subject will help them to understand the basic principles of the subject more easily, improve the effect of memory and promote learning transfer.

The knowledge points of high school subjects are scattered and complicated, but the connection between knowledge points is strong and the flexibility is high. According to the characteristics of high school knowledge and cognitive structure learning theory, this system introduces knowledge graph technology to show the structure of course knowledge for students [19], and promote the construction of students' meaning of knowledge.

4.2 Design of teaching application mode

According to SAM model, teaching design can be divided into preparation stage, iterative design stage and iterative

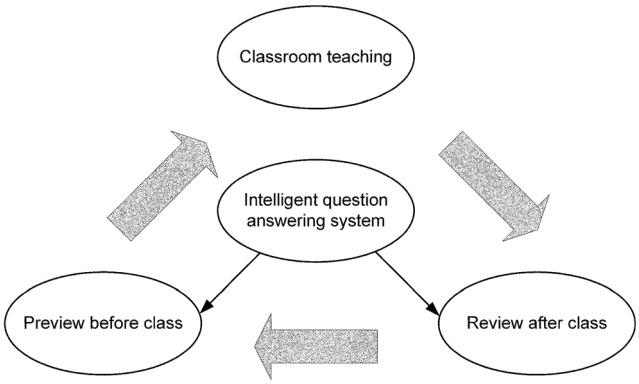

Fig. 4 Teaching application mode

development stage. This paper designs the teaching application mode of the system based on SAM model, as shown in Fig. 4.

In the preview before class, students use the intelligent q&a system to answer questions, and can understand the structure of knowledge points in advance; In classroom teaching, the intelligent q&a system collects and analyzes the data of students' questions, and returns the analysis results to the teacher, who designs the teaching according to the results; In the after-class review, students use the intelligent q&a system to check the gaps and make up the gaps, and consolidate what they have learned in class.

In this model, the intelligent q&a system based on the knowledge graph of high school course connects pre-class preview, classroom teaching and after-class review to form a closed loop. In the preparation stage, the teachers collect the student questions information through intelligent q&a system, so as to understand students' weak spots, then carry on the teaching design according to the students learning situation, and next to classroom teaching. After this loop is completed, the next loop is repeated, in order to form a complete teaching feedback mechanism.

5 Conclusion

This paper firstly analyzes the problems existing in the question answering of high schools, and puts forward that the intelligent q&a system based on knowledge graph is an effective way to solve these problems. Then, under the guidance of constructivism learning theory and cognitive structure learning theory, this paper designs an intelligent q&a system based on high school course knowledge graph by using knowledge graph technology, intelligent q&a technology and big data technology. This system is the innovative achievement of the new generation of information technology in the field of Education. It can not only lighten the work burden of teachers, give teachers feedback on students' learning, but also answer students' academic questions timely and accurately, and help students to build their knowledge structure. Finally, according to the SAM model, this paper designs the teaching application mode of the system. The next step is to use deep learning method to improve the accuracy of problem understanding, expand and enrich knowledge graph base, and strive to expand the application of the system to all levels of education.

Acknowledgments The research is supported by Program for Yunnan Key Laboratory of Smart Education, and Program for innovative research team(in Science and Technology) in University of Yunnan Province.

References

1. Miftachul H, Andino M, Pardimin A, Maragustam S, Roslee A, Jasmi Kamarul A, Nor MNH (2018) Big data emerging technology: insights into innovative environment for online learning resources[J]. Int J Emerg Technol Learn 13(01):23
2. Paul PK, Ghose MK (2018) A novel educational proposal and strategies toward promoting cloud computing, big data, and human–computer interaction in engineering colleges and universities[M]. Advances in Smart Grid and Renewable Energy
3. Knobloch J, Kaltenbach J, Bruegge B (2018) Increasing student engagement in higher education using a context-aware Q&A teaching framework[C]. the 40th International Conference
4. Zhiyun C, Yue S, Dongming Q (2018) Research on intelligent question answering system based on knowledge map [J]. Comput Appl Software 35(02):178–182
5. Cheng C, Jie Z, Jinyu Q, Jia J, Wu H, Tingting C (2018) Research on intelligent question answering technology based on TCM knowledge map [J]. China New Commun 20(02):204–207
6. Chen P, Lu Y, Zheng VW, Xiyang C, Boda Y (2018) KnowEdu: a system to construct knowledge graph for education[J]. IEEE Access PP(99):1–1
7. Wei, QF, Cao, CZ, Wang, FR, Luo, CS (2017) Q&A system development based on agricultural knowledge map and multi-level association model[J]. Adv Soc, Educ Hum 61(−):536–538
8. Liu Y, Xu B, Yang Y, Tonglee C, Zhang P (2019) Constructing a hybrid automatic Q&A system integrating knowledge graph and information retrieval technologies[M]. Foundations and Trends in Smart Learning
9. Ye D, Xing Z, Kapre N (2017) The structure and dynamics of knowledge network in domain-specific Q&A sites: a case study of stack overflow[J]. Empir Softw Eng 22(1):375–406
10. Huang Huan M, Tingting H, Linjing W (2019) A study on the construction of curriculum knowledge map for adaptive learning system – taking "Java programming foundation" as an example [J]. Modern Educ Technol 29(12):89–95
11. Mingyu C, Qingqing L, Yang Z, Wang L, Yin Z, Lin H, Wang J (2019) Knowledge query system of primary liver cancer based on knowledge map [J]. Chin J Inform 33(06):88–93
12. Zhe L, Meng C, Takanori M, Li J (2018) Construction and application of korean-english-japanese multilingual teaching aid system based on knowledge map[J]. Int J Distance Educ Technol 16(4):1–14
13. Peng Z, Wei Z, Xianming Y (2019) Auto-construction of course knowledge graph based on course knowledge[J]. Int J Performability Eng 15(8):2228
14. Sharma Y, Gupta S (2018) Deep learning approaches for question answering system[J]. Procedia Computerence 132:785–794
15. Sun G, Chen T, Su Y, Li C (2018) Internet traffic classification based on incremental support vector machines[J]. Mobile Networks Appl 23(4):789–796

16. Sun G, Li J, Dai J, Song Z, Lang F (2018) Feature selection for IoT based on maximal information coefficient[J]. Future Gen Comput Syst-Int J Esci 89:606–616
17. Wen J, Zhang W, Shu W (2018) A cognitive learning model in distance education of higher education institutions based on chaos optimization in big data environment[J]. J Supercomput
18. Zhang W, Jiang L (2018) Algorithm analysis for big data in education based on depth learning[J]. Wirel Pers Commun
19. Qiao L, Yang L, Hong D, Yao L, Zhiguang Q (2016) Overview of knowledge map construction technology [J]. Comput Res Develop 53(03):582–600

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.