

## КОМПЬЮТЕРНАЯ СИСТЕМА TAXON.PRO ДЛЯ ИДЕНТИФИКАЦИИ И ИССЛЕДОВАНИЯ БИОЛОГИЧЕСКИХ ТАКСОНОВ В СЕТИ ИНТЕРНЕТ

А.Г. Хмарик<sup>1</sup>, Д.Д. Сластунов<sup>1</sup>, А.А. Егоров<sup>1,2</sup>, А.К. Сытин<sup>3</sup>,  
Л.В. Орлова<sup>3</sup>, А.Л. Лобанов<sup>4</sup>

<sup>1</sup>Санкт-Петербургский государственный лесотехнический университет,  
194021, Санкт-Петербург, Институтский пер., 5, e-mail: info@taxon.pro

<sup>2</sup>Санкт-Петербургский государственный университет, 199034, Санкт-Петербург, Университетская наб., д. 7–9

<sup>3</sup>Ботанический институт имени В.Л. Комарова РАН, 197376, Санкт-Петербург, ул. Профессора Попова, д. 2

<sup>4</sup>Зоологический институт РАН, 199034, Санкт-Петербург, Университетская наб., д. 1

Рассмотрены особенности компьютерной системы для описания, исследования и определения различных биологических таксонов по морфологическим признакам, а также освещены проблемы перевода информации в электронный формат. Описаны примеры создаваемых баз данных – “Аборигенные и интродуцированные таксоны хвойных растений Северо-Запада европейской части России” и “Дендрофильные насекомые средней полосы России”, а также некоторые другие отечественные и зарубежные электронные базы данных.

**Ключевые слова:** биологическая систематика, биодиагностика, информационные технологии, компьютерная обработка данных, многоходовые определители.

## WEB SOFTWARE PACKAGE FOR IDENTIFYING AND RESEARCHING BIOLOGICAL TAXONS

A.G. Khmarik<sup>1</sup>, D.D. Slastunov<sup>1</sup>, A.A. Egorov<sup>1,2</sup>, A.K. Sytin<sup>3</sup>,  
L.V. Orlova<sup>3</sup>, A.L. Lobanov<sup>4</sup>

<sup>1</sup>St. Petersburg Forest Technical University,  
194021, St. Petersburg, Institutsky Lane, 5, e-mail: info@taxon.pro

<sup>2</sup>St. Petersburg State University, 199034, St. Petersburg, Universitetskaya nab., 7–9

<sup>3</sup>Komarov Botanical Institute RAS, 197376, St. Petersburg, Prof. Popov str., 2

<sup>4</sup>Zoological Institute RAS, 199034, St. Petersburg, Universitetskaya nab., 2

The article describes the features of computer system for description, studying and identification of various biological taxa, based on a morphological characters. It highlights problems of translation of information into an electronic format. Described examples of created databases – “Native and introduced conifers taxa of North-West European part of Russia” and “Dendrophilous insects of Central Russia”, as well as some other domestic and foreign electronic databases.

**Key words:** biological systematics, biodiagnostics, information technology, computer data processing, multi-input determinants.

## ВВЕДЕНИЕ

Одним из существенных моментов научной деятельности и образования является доступность информации. Объем данных, накопленный за годы научной работы, огромен, но большая его часть хранится в виде опубликованных монографий, специализированных сводок или статей в периодических изданиях, а также в виде рукописей. Часто (из-за относительной малодоступности и удаленности библиотек и личных архивов) значительная часть информации лежит “мертвым грузом” и практически не используется, несмотря на ее научную значимость. Печатные носители не позволяют производить быструю систематизацию и анализ

возросшего массива данных из разных источников. В итоге специалисты часто не успевают обрабатывать информацию в полном объеме.

Улучшить эффективность научной работы в области биологии, и, в частности, в систематике растений, поможет использование современных компьютерных технологий. Поэтому наша цель – создать компьютерную систему, способную повысить уровень преимущества научной информации в биологии и систематике растений. Таким образом, предполагается ввести в научный оборот значительную часть прежде не использованного материала. Работа по созданию программного

комплекса под названием “БИО-информационные системы Taxon.pro” проводится нами с 2010 г. на базе Санкт-Петербургского государственного лесотехнического университета (кафедра ботаники и дендрологии, а также кафедра защиты леса и охотоведения) совместно с сотрудниками Ботанического института РАН (отдел Гербарий высших растений) и Зоологического института РАН (лаборатория систематики насекомых).

Первые опыты использования электронных определителей биологических объектов были проведены в середине 1960-х годов (Кискин и др., 1965; Ladley, 1965; Goodall, 1968; Morse, 1968; и др.). На базе ЗИН РАН А.Л. Лобанов разрабатывал теоретические и практические основы создания электронных определителей с 70-х годов XX века (Лобанов и др., 2006). Созданная им на базе ЭВМ “Наири-1” в 1974 г. диагностическая система “Диагностика-1” стала первой в СССР. В 2002 г. коллективом авторов совместно с А.Л. Лобановым (Смирнов и др., 2010) была разработана информационно-поисковая система “Биоразнообразие России” – БИОДИВ (BIODIV). Она была размещена на веб-портале Зоологического института РАН. В 2010 г. при содействии А.Л. Лобанова была опубликована компьютерная программа SUPERKEY (Кирейчук и др., 2011), позволяющая производить

многовходовое определение в сети Интернет (<http://www.zin.ru/projects/webkey-x/superkey/>). На базе этой программы функционируют два многовходовых политомических определителя – для 130 видов отряда *Coleoptera* и 35 видов класса *Ophiuroidea*.

В 2010 г. в лаборатории Гербарий ЦСБС СО РАН И.А. Артемовым создана компьютерная программа “Многовходовый политомический компьютерный ключ для определения астрагалов Сибири” (Артемов, 2010). Программа позволяет проводить многовходовое определение 102 видов и подвидов рода *Astragalus* L. с возможностью расчета диагностической ценности признаков.

В 2012 г. Т.А. Остроумовой (Ботанический сад МГУ) создан электронный определитель зонтичных (*Umbelliferae*), который опубликован на CD-ROM-диске как приложение к монографии “Зонтичные (*Umbelliferae*) России” (Пименов, Остроумова, 2012). Определитель содержит 300 таксонов *Umbelliferae* и позволяет проводить их идентификацию по 81 признаку.

В 2004 г. был создан веб-ресурс “Plantarium” – атлас видов с первым онлайн-определителем растений. Основной задачей ресурса была систематизация данных и обмен опытом между специалистами в области ботаники и любителями природы.

## МАТЕРИАЛ И МЕТОДЫ

Алгоритм проделанной нами научно-исследовательской работы при составлении базы данных (БД) для многовходового иллюстрированного политомического ключа можно разбить на следующие этапы:

1) формирование коллективов специалистов, включающих экспертов в области различных таксономических групп;

2) изучение литературных источников по систематике и морфологии исследуемой группы таксонов, в том числе первоописаний различных таксонов. Параллельно осуществляется формализация признаков биологических объектов, уточняется терминология и производится отбор наиболее таксономически значимых диагностических признаков;

3) установление достоверности данных о морфологических признаках таксонов на основе сравнения с их первоописанием и, по возможности, с типовыми и другими гербарными образцами в различных отечественных и зарубежных гербариях, а также с особями в природных популяциях и дендроколлекциях;

4) выполнение детальных макроснимков вегетативных и репродуктивных органов растений, служащих одним из основных объектов таксономической информации;

5) осуществление эмпирической оценки надежности каждого значения и признака в целом. Разработка шкалы ценности признаков и оценка их надежности в многовходовом политомическом ключе выполняются с использованием математического анализа по методу Дж.А. Маккаро (Свиридов, 1994), применяющегося для сокращения средней длины определения в одноходовых определительных ключах.

Аборигенные и интродуцированные хвойные растения Северо-Запада РФ были выбраны в качестве первых исследуемых таксонов. В основу создаваемой БД вошли диагностические признаки вегетативных и репродуктивных органов, уточненные и новые для этой группы растений. Данные о них были получены в ходе многолетних комплексных исследований различных таксонов хвойных (Орлова, 2001, 2012; Фирсов, Орлова, 2008; Орлова и др., 2011). Монографом некоторых групп голосеменных – Л.В. Орловой (БИН РАН) была проделана работа по систематизации диагностической информации. Основные трудности были с установлением единых стандартных значений признаков. Дело в том, что для многовходовых определителей формулировка признаков требует одновременно четкости и ограничения количества состояний признаков от 2 до 9 в соответствии с

так называемым числом Миллера (Miller, 1956). В связи с этим выполнялась формализация признаков, учитывающая максимально четкую (однозначную) трактовку их значений. Например, в крупнейших монографиях и определителях по хвойным количество состояний для одного признака в списке превышало 10–20, а иногда и более. В работе Г. Крюссмана “Хвойные породы” (1986) только для характеристики окраски молодых побегов видов рода *Abies* упоминается 16 цветов и их оттенков.

Качественная фотофиксация морфологических структур растения имеет исключительно важное значение для метода нашей работы. Осуществлялись съемка габитуальных признаков (габитус деревьев и кустарников, морфологические особенности коры, расположение ветвей первого и второго порядков), макросъемка вегетативных и репродуктивных органов, а также съемка объектов под бинокулярным микроскопом “Микромед 1 вар. 2–20”. На стереоскопическом люминесцентном микроскопе “SteREO Lumar.V12” сфотографирована поверхность побегов и хвоинок различных таксонов *Larix*. Для фотосъемки материала разработана технология с использованием масштабных линеек и цветокалибровочных шкал (рис. 1, 2) и специальной системы освещения.

**Программный комплекс “Taxon.pro”. Описание структуры компьютерной системы.** В ходе разработки программного комплекса мы опирались на материалы последних 50 лет по применению информационных технологий в биодиагностике. Также нами был использован опыт составления печатных цифровых политомических ключей (Балковский, 1964; Кискин и др., 1965; Бородина и др., 1966). В указанных работах особое внимание уделялось методике построения многовходовых ключей. Подобный тип ключей позволяет определять таксон в разных комбинациях признаков и при любом порядке выбора установленных значений признаков.



Рис. 1. Многолетний побег *Larix olgensis* A. Henry.

Программа для заполнения данных и программа для исследования и идентификации биологических объектов выполнены на платформе ExtJs 4 (Sencha) и позволяют работать как удаленно (веб-приложение на сервере), так и локально (без доступа к Интернету). Ключевой структурой системы является расширяемая база данных, позволяющая описывать биологические объекты по их характеристикам.

Программа для заполнения данных доступна для зарегистрированных и верифицированных на сайте <http://bis.taxon.pro/> авторов. Программа для исследования и идентификации таксонов доступна всем пользователям Интернета по адресу: <http://bis.taxon.pro/> и позволяет использовать открытые БД без необходимости регистрации.

**Базу данных** можно представить как таблицу состояний признаков, непосредственно относящихся к таксонам. Эта виртуальная таблица является многомерной и поэтому поддерживает поливариантность состояний признака у конкретного таксона. К каждому значению в таблице можно добавить комментарии и другие метаданные, а также ссылки на библиографические источники и оригинальные фотографии.

**Программа для заполнения данных** имеет существенные достоинства для ввода сведений специалистами-биологами. Прежде всего это широкая доступность и простота их внесения. Возможность использования как в сети Интернет, так и локально. Отсутствие ограничения на объем информации и возможность усложнения структуры базы данных. Возможность одновременной сетевой работы нескольких авторов и наличие систематизированной связи между ними.

**Программа для исследования и идентификации таксонов.** Конечный пользователь этой программы имеет возможность определять растение с помощью многовходового определителя, исследовать содержащиеся в базе таксоны, используя визуальную информацию, сравнивать группы так-



Рис. 2. Чешуи шишек *Abies mariesii* Mast.

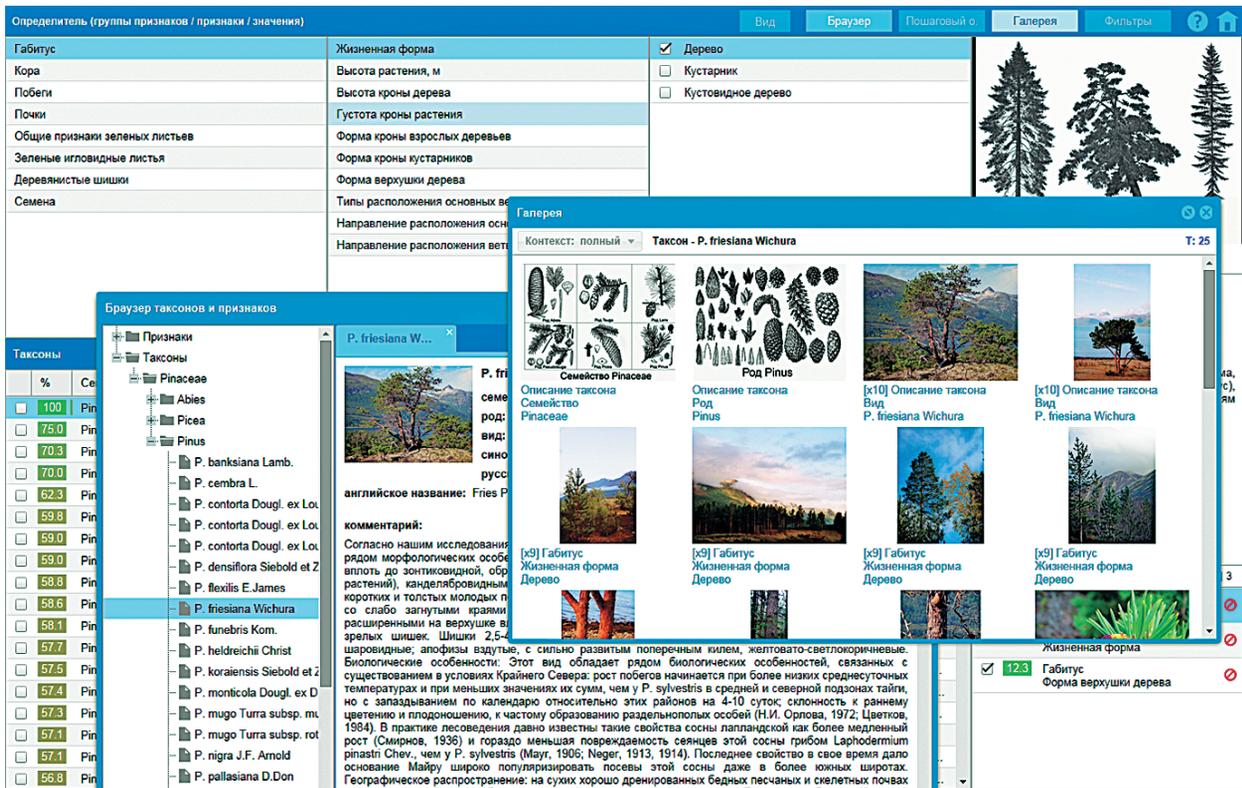


Рис. 3. Интерфейс программы Taxon.pro для идентификации и исследования таксонов в сети Интернет.

сонов или группы признаков с применением алгоритмов математической обработки данных. В режиме пошагового определителя производится автоматический расчет наиболее подходящих для определения признаков и вычисление надежности идентификации объекта на каждом шаге определения. В приложении предусматривается возможность сравнивать фотографии и рисунки, иллюстрирующие систематические признаки таксонов, что повышает точность их идентификации (рис. 3).

В ближайшем будущем планируется введение поддержки построения различных графиков

и диаграмм, иллюстрирующих связи признаков и таксонов внутри выбранных для изучения групп. В связи с этим можно проводить анализ различных дендрограмм, полученных в результате морфологических исследований, тем самым позволяя уточнять морфологические характеристики биологических таксонов, а при использовании данных молекулярно-генетических исследований и их филогенетические связи. Также планируется внедрение прохождения учебных тестов на знание свойств, описанных в базах данных таксонов.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

К концу 2014 г. нами были достигнуты следующие результаты:

- созданы и апробированы программные продукты для заполнения данных и исследования таксонов. Методы работы с данными, используемые в этих средствах, прошли экспертную проверку специалистов по морфологии и систематике растений;

- разработаны и апробированы методы качественной фотофиксации морфологических признаков биологических объектов. Изготовлены установки специализированного предметного освещения, написаны программные продукты для каталогизации отснятого фотоматериала;

- разработаны методы качественной фотосъемки в полевых условиях.

Заполнены следующие БД, включая внесение иллюстраций и комментариев.

1. Аборигенные и интродуцированные виды хвойных растений Северо-Запада России, определитель и атлас (автор базы данных Л.В. Орлова). База содержит 137 таксонов и охватывает максимально возможное количество устойчивых морфологических признаков (130 признаков, объединенных в 12 групп), многие из них оригинальные. Источником данных послужили многолетние комплексные исследования хвойных во многих отечественных и зарубежных гербариях (LE, LECB, KFTA, MW, VLA, BP, PE, KRAM, KOR, H, O и др.), а также в природе – на Дальнем Востоке (п-ов Камчатка и Южное Приморье), в Восточной (Якутия, Прибайкалье, окрестности Красноярска) и Запад-

ной Сибири (окрестности Новосибирска, п-ов Ямал), а также на территории европейской части России (Архангельская, Мурманская, Новгородская, Псковская области, Карелия, Краснодарский и Ставропольский края, Крым), ближнего зарубежья (Украина, Республика Кыргызстан) и Западной Европы (Польша, Швейцария), и ботанических коллекциях (Орлова, 2001, 2012; Orlova, 2005; Фирсов, Орлова, 2008).

2. Дендрофильные насекомые средней полосы России: определитель и атлас (авторы базы данных Л.Н. Щербакова, Н.В. Денисова (Хмарик и др., 2014)). База данных и многоходовый ключ для определения видовой принадлежности вредителей древесных растений по оставляемым ими повреждениям. База ориентирована на работников лесного и садово-паркового хозяйства, владельцев дачных участков и любителей живой природы, а также студентов вузов. В нее вошли 255 видов дендрофильных насекомых, отмеченных в Санкт-Петербурге и Ленинградской области. Это аборигенные и интродуцированные представители 39 семейств, входящих в 7 отрядов (Гусев, Римский-Корсаков, 1934).

В созданных за это время БД были уточнены списки морфологических признаков и диапазоны их значений для различных таксонов (Хмарик и др., 2013). В процессе работы усовершенствован подход к созданию БД многоходовых политомиических ключей.

Разработанный программный комплекс восполняет отсутствие на сегодня систем, сочетающих в себе инструменты создания БД многоходовых диагностических ключей, а также интерактивные средства анализа текстовой и графической информации для идентификации таксонов.

Внедренные нами инструменты позволяют разрабатывать эффективные многоходовые определители, а также делают информацию более точной и доступной для проверки, что было доказано в ходе испытания созданного в рамках проекта многоходового диагностического ключа аборигенных и интродуцированных таксонов хвойных растений Северо-Запада европейской части России. Указанный диагностический ключ позволяет достоверно определять растения, относящиеся ко многим видам, даже при отсутствии у образцов каких-либо органов.

Такой подход обеспечивает для авторов БД ряд преимуществ по сравнению с традиционными способами обработки и публикации материалов. В первую очередь – это упорядочивание большого объема информации, накопленной автором. Кроме того, размещение данных в сети Интернет снимает ограничения на объем публикуемого матери-

Опыт применения системы выявил необходимость совершенствования систематизации данных и формализации терминов описательной морфологии в процессе переноса данных в электронный формат. При проведении этой работы многие данные могут быть дополнительно уточнены. Наличие большого количества иллюстрирующих фотографий также позволяет провести дополнительную проверку точности вводимых данных.

**Сравнение с международным опытом.** Отметим, что направление компьютерной систематизации биологических данных развивается довольно слабо как в нашей стране, так и за рубежом. Ни одна из существующих электронных систем полностью не удовлетворяет современным требованиям работы с информацией.

Системы, подобные Delta (<http://delta-intkey.com/>) и LucID (<http://www.lucidcentral.com/>), не предоставляют возможности статистического анализа взаимосвязей таксонов и не содержат в себе специализированных алгоритмов, упрощающих работу пользователя. Также данные проекты не имеют простого онлайн-доступа.

На настоящий момент в системе “БИО-информационные системы Taxon.pro”, находящейся по адресу <http://bis.taxon.pro/>, в открытом доступе опубликована часть базы данных “Аборигенные и интродуцированные виды хвойных растений Северо-Запада России” (роды *Abies*, *Picea* и *Pinus*).

## ЗАКЛЮЧЕНИЕ

ала, в том числе графического, а индексируемость авторской информации в поисковых системах расширяет круг пользователей и, как следствие, цитируемость материала.

Описанные выше подходы повышают эффективность работы систематиков, а также делают идентификацию таксонов, как и научное исследование их характеристик, более точной. Следует отметить, что внедрение компьютерных баз данных, рассчитанных на использование в многоходовых диагностических ключах, является перспективным направлением и требует комплексного подхода широкого круга специалистов (морфологов, систематиков, генетиков). Дальнейшие исследования предполагают испытание нашего программного комплекса на базе других таксономических групп растений, в частности модельной группы однолетних астрагалов (*Astragalus* L., *Fabaceae*).

*Работа выполнена при поддержке Минобрнауки РФ в рамках научного проекта № 2014/181-2220, Департамента по науке и инновациям Ямало-Ненецкого автономного округа по госконтракту от 25 июля 2012 г. № 01-15/4, а также гранта РФФИ № 15-04-06981 А.*

## ЛИТЕРАТУРА

- Артемов И.А.** Многоходовый политомический компьютерный ключ для определения астрагалов Сибири (*Astragalus L., Fabaceae*) // Сиб. экол. журн. 2010. Т. 17, № 6. С. 909–918.
- Балковский Б.Е.** Цифровой политомический ключ для определения растений. Киев, 1964. 36 с.
- Бородина Н.А., Некрасов В.И., Некрасова Н.С.** Деревья и кустарники СССР. М., 1966. 637 с.
- Гусев В.И., Римский-Корсаков М.Н.** Определитель повреждений лесных и декоративных деревьев и кустарников европейской части СССР. Л., 1934. 430 с.
- Кирейчук А.Г., Лобанов А.Л., Смирнов И.С., Иночкин А.А., Степаньянц С.Д.** Интернет-определители биологических объектов. 5 лет спустя // Международный суперкомпьютерная конф. “Научный сервис в сети Интернет: экзотическое будущее” (19–24 сент. 2011 г., Новороссийск). М., 2011. С. 449–453.
- Кискин П.Х., Печерская И.Н., Печерский Ю.Н.** Автоматизация диагностического поиска сортов винограда на ЭВМ “Минск-1” // Виноделие и виноградарство СССР. 1965. № 1. С. 21–22.
- Крюссман Г.** Хвойные породы. М., 1986. 256 с.
- Лобанов А.Л., Кирейчук А.Г., Смирнов И.С., Граничин О.Н., Вахитов А.Т., Дианов М.Б.** К реализации идеального интерактивного определителя биологических объектов в Интернете // Тр. Всерос. науч. конф. “Научный сервис в сети Интернет: технологии параллельных вычислений” (18–23 сент. 2006 г., Новороссийск). Новороссийск, 2006. С. 202–204.
- Орлова Л.В.** Систематический обзор дикорастущих и некоторых интродуцированных видов рода *Pinus L. (Pinaceae)* флоры России // Новости сист. высш. раст. Л., 2001. Т. 33. С. 7–40.
- Орлова Л.В.** *Pinophyta* – Голосеменные // Конспект флоры Восточной Европы. Т. 1. М., 2012. С. 49–90.
- Орлова Л.В., Фирсов Г.А., Егоров А.А., Неверовский В.Ю.** Хвойные Санкт-Петербургской лесотехнической академии: Аннотированный каталог. СПб., 2011. 88 с.
- Пименов М.Г., Остроумова Т.А.** Зонтичные (*Umbelliferae*) России. М., 2012. 477 с. + 1 эл. опт. диск (CD-ROM).
- Свиридов А.В.** Ключи в биологической систематике: теория и практика. М., 1994. 224 с.
- Смирнов И.С., Пугачев О.Н., Кирейчук А.Г., Дианов М.Б., Лобанов А.Л., Халиков Р.Г., Голиков А.А., Кривохатский В.А.** Итоги и перспективы разработки информационной системы по биоразнообразию животных России (ZOODIV – BIODIV) // Двенадцатая Всерос. конф. RCDL на базе Казанского (Приволжского) федерального университета (13–17 окт. 2010 г., Казань). Казань, 2010. С. 461–464.
- Фирсов Г.А., Орлова Л.В.** Хвойные в Санкт-Петербурге. СПб., 2008. 336 с.
- Хмарик А.Г., Сластунов Д.Д., Васильев Н.П., Орлова Л.В., Егоров А.А., Иванов С.А., Бялт А.В.** Разработка электронного определителя хвойных северо-запада России // Современная ботаника в России: Тр. XIII Съезда Рус. ботан. о-ва и конф. “Научные основы охраны и рационального использования растительного покрова Волжского бассейна” (16–22 сент. 2013 г., Тольятти). Тольятти, 2013. Т. 2. С. 80–81.
- Хмарик А.Г., Сластунов Д.Д., Денисова Н.В., Щербакова Л.Н.** Многоходовый электронный определитель дендрофильных насекомых // VIII Чтения памяти О.А. Катаева. Вредители и болезни древесных растений России: Материалы Международной конф. (18–20 нояб. 2014 г., Санкт-Петербург) / Под ред. Д.Л. Мусолина, А.В. Селиховкина. СПб., 2014. С. 87.
- Goodall D.W.** Identification by computer // BioScience. 1968. V. 18, No. 6. P. 485–488.
- Ladley R.S.** Use of computer in biology and medicine. N.Y., 1965. 965 p.
- Miller G.A.** The Magical Number Seven, Plus or Minus Two // Psychol. Rev. 1956. V. 63. P. 81–97.
- Morse L.E.** Construction of identification keys by computer // Amer. J. Botan. 1968. V. 55, No. 6. P. 737.
- Orlova L.V.** On the systematic importance of the morphology and anatomy of the vegetative and reproductive organs in the *Pinaceae* // Materials of XVII International Botanical Congress. Vienna, 2005. P. 386.