

# Информационный анализ признаков растений

**В.И. Авдеев**, д.с.-х.н., профессор,  
ФГБОУ ВО Оренбургский ГАУ

В современной теоретической и прикладной биологии всё шире используют методы, связанные с биотермодинамикой, информатикой равновесных (белковые структуры видов, энергетика фотосинтеза) и неравновесных состояний биосистем (биологическая самоорганизация). Как считают, сейчас без теории информации не может идти и речи о современной биологии [1]. Известно, что информатика как наука формировалась на основе теорий вероятностей и термодинамики [2, 3]. По К.Э. Шеннону [2], информация является мерой неопределённости (или энтропии), исчезающей при получении сведений о системе. Применительно к биологии даны и другие определения информации: это сообщение, сигнал, мера многообразия в строении популяций, результат неупорядоченности в них и т.д. [4]. За единицу информации был принят бит (*binary digits* – англ.), т.е. количество информации в достоверном сообщении о событии, априорная вероятность которого равна 1/2. В этой связи используют двоичный код («да» или «нет»), а в формуле информации – логарифм с основанием 2. В биологии события ( $p_1 \dots p_n$ ) – неравновероятные, при этом рассчитанную информацию называют как «энтропия сообщения ( $H$ )», или же «энтропией множества вероятностей  $p_1 \dots p_n$ », которую выражают с помощью известной формулы К.Э. Шеннона [2, 5 и др.]. Но отметим, поскольку  $H$  названа энтропией, то, как сказал Б.М. Медников [5], это внесло «немало сумятицы в умы биологов». На самом деле  $H$  является формальным аналогом термодинамической энтропии ( $S$ ). Более того,  $H$ , правда, разово на мизерную величину, но

снижает величину  $S$  приёмника информации, т.е. ведёт себя как внешняя энергия.

При осмыслении значений  $H$  также выявляются противоречия. Так, Н.А. Плохинский [4] считал, что в биологии нужно учитывать максимальные значения  $H$ , а Г.Н. Зайцев [6] рекомендовал в ботанике учитывать только её минимальные значения. Последнее же, несомненно, связано с тем, что  $H$  полагали как часть  $S$ , поскольку известно, что в открытых неравновесных биосистемах энтропия падает ( $S \rightarrow \min$ ), что приводит систему к стационарному равновесию, упорядоченности её структур [7 и др.]. Также необходимо знать, как изменяются значения  $H$  при разном статистическом распределении значений (дат, событий) количественных признаков: по Гауссу (нормальном), асимметричном ( $As$ ), эксцессе ( $E$ ). **Задача статьи** – на конкретных примерах по обычным признакам растений показать применимость теории информатики, рассмотреть названные выше проблемы.

**Материал и методы исследования.** Для наглядности часть данных взята из публикаций автора статьи [8–11]. Методы исследования и биостатистической обработки данных приведены в тех же работах. Обозначали  $M$  – среднее значение признака;  $P$  – уровень значимости различий по признакам;  $n$  – число особей (шт.);  $k$  – число классов в распределении признаков (шт.). Расчёт же  $H$  проводили по принятой схеме [6], но использовали вспомогательную таблицу [4].

**Результаты исследования.** Пример (табл. 1) важен тем, что при изучаемом распределении признака – масса косточки (эндокарпия) у миндаля обыкновенного (*Amygdalus communis* L.) – в популяции найдена правосторонняя (положительная) асимметрия признака. Так, правее класса максимальных значений признака (2,20–2,51;  $M=2,50$  г) находится 64, а левее – лишь 49 частот признака. Коэффициент асимметрии  $As=+0,46$  (при  $0,05 > P > 0,01$ ). Эта асимметрия указывает на то, что в популяции имеет место движущая форма естественного отбора [10].

Дальнейшие расчёты сведены в таблицу 2. Так, для сравнения приводим близкий пример из популяции миндаля туркменского (*A. turcomanica* Lincz.), где по распределению признака в пределах класса 0,23–0,32 г был выявлен эксцесс (островершинность), т.е.  $E=+1,03$  (при  $P < 0,001$ ). Но такое распределение уже свойственно стабилизирующему естественному отбору, когда преимущественно выживают средние по признаку особи [10]. Из этих же данных видно, что в случае объединении сведений по двум популяциям миндаля бухарского – *A. bucharica* Korsh. (число учётных особей стало  $n=125$  шт.) значения  $H$  выросли в среднем на 14% ( $H$  стал от 2,20 до 2,62). Другой пример – распределение массы косточек у микровишни бородавчатой [*Microcerasus prostrata* var. *verrucosa* (Franch.) Erem.

1. Сырая масса косточки ( $x_i$ ) миндаля обыкновенного из природных условий Юго-Западного Копетдага

Исходные данные значений признака			Частности, $p_i=f/n$	$H$ , в битах
границы классов, г ( $x_i$ )	середина класса, г ( $x_i$ )	частота, шт. ( $f$ )		
0,92–1,23	1,08	3	0,0207	0,111
1,24–1,55	1,40	3	0,0207	0,111
1,56–1,87	1,70	14	0,0966	0,312
1,88–2,19	2,04	29	0,2000	0,460
2,20–2,51	2,36	32	0,2206	0,482
2,52–2,83	2,68	26	0,1793	0,438
2,84–3,15	3,00	17	0,1172	0,350
3,16–3,47	3,32	10	0,0690	0,243
3,48–3,79	3,64	6	0,0414	0,190
3,80–4,11	3,96	2	0,0138	0,072
4,12–4,43	4,28	3	0,0207	0,111
$c = 0,31$	$k = 11$	$n = 145$	$\Sigma = 1,0000$	$\Sigma = 2,88$

et Yushev]. Здесь нет как асимметрия, так и эксцесса [9], распределение достоверно не отличается от закона нормального распределения Гаусса ( $P > 0,05$ ). Эти же данные были рассредоточены на большее число классов,  $k = 11$  [11], что привело почти на 8% к росту значений  $H$ : от 2,74 (при  $k = 9$ ) до 2,97 (табл. 2).

Близкие значения  $H$  получены по массе плода в популяциях ореха грецкого (*Juglans regia* L.) и яблони туркменов (*Malus turkmenorum* M. Pop. et Juz.), и что важно, это почти не связано с разницей по  $n$  – числу изученных особей (табл. 2).

Близкими оказались данные по высоте растений у облепихи (*Hippophae rhamnoides* L.). Что касается довольно низкого значения  $H = 1,43$  у боярышника туркестанского (*Crataegus turkestanica* Pojark.), то это объясняется лишь слабым варьированием у него признака «высота дерева» (от 2,3 до 5,5 м), что в 1,5–2 раза ниже, чем по признакам других видов. Иначе говоря, то, что  $H \rightarrow \min$ , возникает только из-за однообразности (низкой информативности) этого признака. По этой же причине различными (в 1,3 раза) получены значения  $H$ , появившиеся между популяциями барбариса разнокистевидного (*Berberis heterobotrys* E. Wolf). Но это связано с тем, что в популяции «Шахтут» человек вырубал для лекарственных целей сильнорослые кусты, так что высота оставшихся особей оказалась от 1,5 до 2,5 м [11], т.е. здесь в итоге возникло несколько большее однообразие признака,  $H = 1,91$  (табл. 3).

Обычные значения  $H$  (от 2,48 до 2,82) получены и при изучении сортов, форм абрикоса (*Armeniaca vulgaris* Lam.) в культивируемых популяциях (табл. 4). В работе Г.Н. Зайцева [6] близкие данные (в пересчёте на биты) установлены при изучении фенологических фаз растений-интродуцентов. Этот автор был неточен, учитывая лишь низкие значения  $H$ , т.е. те, которые присущи близким видам.

Расчёт значений  $H$  часто проводят при анализе качественных признаков [4]. Эти значения намного ниже (от 0,40 до 1,38), так как учитывали

три имеющиеся градации признака (табл. 5), а не 11 классов количественных признаков (табл. 1).

Отметим, что у боярышника понтийского (*Crataegus pontica* C. Koch) самые низкие значения  $H$  (0,40 и 0,64) получены там, где в популяциях преобладают соответственно на 88% особи с воронкой у плода (также и две другие градации признака) и 92% (также особи с метамерной изменчивостью по строению воронки у плода). Самое высокое здесь значение  $H = 1,42$  было в популяции, где особей с воронкой у плода было 56%, остальных типов воронки – по 20–24% [9]. То есть падение  $H$  опять-таки связано с низким разнообразием по изучаемому признаку.

Сказанное выше подтверждается расчётами  $H$  на видах миндаля арабский (*A. arabica* Olivier), бухарский, туркменский. Если в популяции преобладает лишь какой-либо фенотип (окраска пестика цветка зелёная или розово-красная), то соответственно это представляет собой падение разнообразия и информативности по этому признаку ( $H$  от 0,42 до 0,49). Отсюда наивысшие значения ( $H = 0,98$ ; практически это значение по К.Э. Шеннону [2]) возможно получить тогда, когда соотношение фенотипов примерно равно, в данном случае – от 41 до 59% (табл. 5). Такое шенноновское значение  $H$  трактуется как исключительно достоверное сообщение. Оно указывает на возникновение в природе состояния стационарного равновесия биосистемы по данному признаку. Известно, что такое её состояние является весьма важным для изучения термодинамических и информационных процессов в биосистемах [1, 3 и др.].

В начале статьи было указано на проблему: учитывать ли в биоинформатике максимальные [4] или же минимальные [6] значения  $H$ ? Далее было выявлено, что большее разнообразие признака даёт повышенные значения  $H$ , меньшее или же однообразие по признаку, наоборот, – наиболее низкое. Тот или иной уровень  $H$  поэтому не может быть предпочтительным для биоинформатики, всё

## 2. Информативность сырой массы плода в природных популяциях

Название и биометрическая характеристика изучаемого объекта	$H$ , в битах
Масса 1 косточки ( $x_i$ ) миндаля обыкновенного из Юго-Западного Копетдага, Западный Туркменистан ( $n = 145$ ; $k = 11$ ; $A_s = +0,46$ при $0,05 > P > 0,01$ )	2,88
Масса 1 косточки ( $x_i$ ) миндаля туркменского из Центрального Копетдага, Туркменистан ( $n = 104$ ; $k = 8$ ; $E = +1,03$ при $P < 0,001$ )	2,61
Масса косточки ( $x_i$ ) миндаля бухарского из Центрального Таджикистана, популяция «Чекурак-1» ( $n = 71$ ; $k = 7$ ; $P > 0,05$ )	2,41
То же, соседняя популяция «Дари-Ишон» ( $n = 54$ шт.; $k = 7$ ; $P > 0,05$ )	2,20
То же, объединённые данные по названным популяциям ( $n = 125$ ; $k = 7$ ; $P > 0,05$ )	2,62
Масса 10 косточек ( $x_i$ ) микровишни бородавчатой из Центрального Таджикистана ( $n = 107$ ; $k = 9$ ; $P > 0,05$ )	2,74
То же ( $k = 11$ )	2,97
То же, соседняя популяция ( $n = 51$ ; $k = 7$ ; $P > 0,05$ )	2,32
Масса 1 ореха (без кожицы) у ореха грецкого из Юго-Западного Копетдага, заповедное ущелье Айidere ( $n = 140$ ; $k = 10$ ; $P > 0,05$ )	2,42
Масса 1 плода яблони туркменов из Юго-Западного Копетдага, массив Алмали-Сеутли ( $n = 32$ ; $k = 6$ ; $P > 0,05$ )	2,20

## 3. Информативность высоты растений в природных популяциях

Название и биометрическая характеристика изучаемого объекта	$H$ , в битах
Высота материнской оси куртины облепихи крушиновидной, Таджикистан, Западный Памир, по реке Ванч, близ кишлака Сутварг ( $n = 104$ ; $k = 11$ ; $P > 0,05$ )	2,97
То же, Дарваз, по реке Висхарви ( $n = 91$ ; $k = 9$ ; $P > 0,05$ )	2,60
То же, Центральный Таджикистан, по реке Оби-Хингоу, близ кишлака Сафедхок ( $n = 60$ ; $k = 8$ ; $P > 0,05$ )	2,22
Высота дерева боярышника туркестанского, Центральный Таджикистан, близ пгт Файзабад и кишлака Новабад ( $n = 50$ ; $k = 4$ ; $P > 0,05$ )	1,43
Высота куста барбариса разнокистевидного, Центральный Таджикистан, склоны левобережья реки Вахш, близ кишлака Кишрог ( $n = 50$ ; $k = 6$ ; $P > 0,05$ )	2,49
То же, склоны правобережья реки Вахш, близ кишлака Шахтут ( $n = 60$ ; $k = 4$ ; $P > 0,05$ )	1,91

## 4. Информативность признаков в культивируемых популяциях

Название и биометрическая характеристика изучаемого объекта	$H$ , в битах
Масса 1 плода у местных абрикосов Ходжакалинской долины, Центральный Туркменистан ( $n = 134$ ; $k = 10$ ; $P > 0,05$ )	2,50
Длина лепестка цветка, там же ( $n = 55$ ; $k = 7$ ; $P > 0,05$ )	2,48
То же, коллекция сортов и форм абрикосов Средней Азии в условиях Юго-Западного Туркменистана, пгт Кара-Кала ( $n = 91$ ; $k = 8$ ; $P > 0,05$ )	2,82

## 5. Информативность природных популяций по качественным признакам

Название и биометрическая характеристика изучаемого объекта	$H$ , в битах
Миндаль бухарский, Центральный Таджикистан, близ кишлака Чукурак-1, доля особей по окраске пестика цветка (красная – 4%, розовая – 4%, зелёная – 92%; $n = 100$ шт.)	0,49
То же, по характеру антоциановой окраски лепестков цветка (лепестки целиком – 15%, лишь жилки – 65%, окраски нет – 20%; $n = 180$ )	1,27
Боярышник понтийский, там же, характер воронки плода в популяциях (она у плода есть или нет, т.е. метамерия в кроне по данному признаку), близ кишлака Новабад ( $n = 100$ шт.)	0,40
То же, близ кишлака Рогун ( $n = 100$ шт.)	0,64
То же, близ кишлака Кишрог ( $n = 100$ шт.)	1,42
Облепиха крушиновидная, Таджикистан, Западный Памир, тип кроны:	
долина Ванча (плакучий – 8%, промежуточный – 47%, пряморослый – 45%; $n = 100$ шт.)	1,32
То же, Дарваз, долина Висхарви (7,3; 45,0 и 47,7% соответственно; $n = 109$ шт.)	1,41
То же, соотношение форм по окраске плода: долина в средней части Ванча, близ кишлака Сутварг (красная – 30%, оранжевая – 55%, жёлтая – 15%; $n = 100$ шт.)	1,24
То же, верховья Пянджа (Амударьи), близ кишлака Абхарв (30; 55 и 15% соответственно; $n = 100$ шт.)	1,31
То же, верховья долины Шахдара (20; 65 и 25% соответственно; $n = 100$ шт.)	1,36
То же, низовья долины Гунт (30,2; 62,3 и 17,5% соответственно; $n = 106$ шт.)	1,38
То же, восток Таджикистана, Припамирье, по реке Оби-Хингоу, близ кишлака Сафедхок (красная и жёлтая – по 7,1%, оранжевая – 85,8%; $n = 42$ шт.)	0,74
То же, север Таджикистана, по реке Ягноб (красная и жёлтая – по 5,4%, оранжевая – 89,2%; $n = 37$ шт.)	0,59
Соотношение фенотипов (особей) по окраске пестика цветка: миндаль обыкновенный, Юго-Западный Копетдаг, по реке Айidere (розово-красная – 8,8% и зелёная – 91,2%; $n = 33$ шт.)	0,42
То же, миндаль арабский, южные склоны Сюнт-Хасардагского хребта (розово-красная – 90,9%, зелёная – 9,9%; $n = 40$ шт.)	0,49
То же, миндаль туркменский, Центральный Копетдаг, местечко Чаек (розово-красная – 41,1%, зелёная – 58,9%; $n = 38$ шт.)	0,98

зависит от задачи исследования. Так, значения  $H \sim 1,0$ , а это есть окраска пестика цветка у миндаля туркменского (табл. 5), совпадает с классическим представлением [2] о равновероятности двух альтернативных событий. Причины же этого – природная антоциановая и хлорофилловая окраска пестика (как итог мутации и естественного отбора, которые возникли в данной популяции миндаля). Заметим ещё раз, что речь идёт не об обмене информацией между биосистемами (здесь учитывают величину  $S$ ),

а об энтропии сообщения, необходимой для наблюдателя (биолога).

Большинство же значений  $H$  по количественным признакам укладывается в пределы от 1,43 до 2,97, по качественным – от 0,40 до 1,42. Это есть показатель многообразия признаков [4], но он вызывается разными внешними причинами. В этой связи есть смысл проанализировать, как влияет на показатель  $H$  отклонение в распределении признаков от нормального, по Гауссу?

Из приведённых данных эти отклонения в основном связаны с асимметрией признака. Это есть данные по сырой массе косточки миндаля на территории Юго-Западного Копетдага (табл. 1), высоте растений облепихи на Дарвазе, высоте дерева боярышника туркестанского в Таджикистане (табл. 3), массе плода и длине лепестка цветка у абрикоса в Ходжакалинской долине (табл. 4). Причина для учёта таких отклонений связана с тем, что при этом растёт разнообразие по признакам, а значит, их значения  $H$ . Но во всех приведённых случаях значения  $H$  изменяются от 1,43 до 2,97, т.е. они не самые большие. Несомненно, это связано с тем, что расчёт  $H$ , согласно методике [2–4 и др.], всегда проводится в логарифмической форме, которая выравнивает существующие в расчётных границах признака отклонения.

**Выводы.** Предпринятые подсчёты на основе экспериментальных данных по изменчивости количественных и качественных признаков в фитосистемах дают ценную информацию для биолога. Значения энтропии сообщения ( $H$ ) составили по объектам исследования 0,40–2,97 бит, что говорит о высоком варьировании этих объектов. Значения  $H$  растут при большом разнообразии признака, т.е. когда идентифицировать объекты трудно, и

падают в случае малого разнообразия, когда объекты можно характеризовать как равновероятные альтернативные события (на примере окраски пестика цветка у миндаля туркменского). Значения  $H$  в случае их расчёта у признаков, распределение которых не подчиняется закону Гаусса (т.е. закону нормального распределения событий), не отличаются от типовых.

### Литература

1. Термодинамика и кинетика биологических процессов. М.: Наука, 1980. 398 с.
2. Шеннон К.Э. Работы по теории информации и кибернетики. М.: ИЛ, 1963. 830 с.
3. Блюменфельд Л.А. Проблемы биологической физики. М.: Наука, 1974. 336 с.
4. Плохинский Н.А. Математические методы в биологии. М.: МГУ, 1978. 268 с.
5. Медников Б.М. Онтогенез и теория информации // Природа. 1971. № 7. С. 15–23.
6. Зайцев Г.Н. Математическая статистика в экспериментальной ботанике. М.: Наука, 1984. 424 с.
7. Пасынский А.Г. Биофизическая химия. М.: Высшая школа, 1968. 432 с.
8. Авдеев В.И., Фаустов В.В., Ермаков Б.С. Биологические особенности облепихи // Облепиха. М.: Лесная промышленность, 1985. С. 12–35.
9. Авдеев В.И. Плодовые растения Средней Азии, их происхождение, классификация, исходный материал для селекции: дис. ... докт. с.-х. наук. СПб.: ВНИИР им. Н.И. Вавилова, 1997. 326 с.
10. Авдеев В.И. Современные методы в исследовании растений. Оренбург: Издат. центр ОГАУ, 2015. 128 с.
11. Авдеев В.И. Изменчивость и биосистематика растений. Оренбург: Издат. центр ОГАУ, 2016. 316 с.