# Limitations of a Random Screen: Search for New Anticancer Drugs in Higher Plants[1]

RICHARD W. SPJUT[2]

*The inherent limitations of a random search of higher plants for novel cancer chemotherapeutic agents are reviewed—the National Cancer Institute's (NCI) Anticancer Screening Program. A graphic summary of plant exploration for the NCI is depicted on a world map showing 58 floristic regions. It is estimated that less than one-half of the world flora is economically feasible for collection. Random screening of approximately 35,000 species has led to guidelines that precluded further screening of all species in 333 genera and another 2,905 species in 1,773 genera. These taxa are reported to represent one-half to two-thirds of the species that characterize vegetation in geographic areas most frequently explored for the NCI. It is estimated that 40,000 untested species of flowering plants are readily available and meet the NCI guidelines for antitumor screening. However, because of apparent diminishing returns from random screening of chemicals in plant genera, it is suggested that a good representation of the diversity in the world flora could be obtained in 10,000 collections, if random sampling follows the phytogeographic outline that is recommended. Modifications to the screening methodology might be geared to an expected point of diminishing returns for discovering novel chemotypes. Additionally, the NCI should continue random screening to increase the development of new anticancer drugs; past screening has generated a tremendous wealth of data. Finally, in this paper, the author proposes to utilize lists representing taxa commonly collected for the NCI to create a manual of worldwide common plants.*

In 25 yr, the National Cancer Institute (NCI) screened more than 120,000 plant extracts from 35,000 species for novel anticancer agents. Some promising discoveries are: taxol, indicine-n-oxide, phyllanthoside, and homoharringtonine, isolated from *Taxus brevifolia* Nutt., *Heliotropium indicum* L., *Phyllanthus acuminatus* Vahl, and *Cephalotaxus harringtonia* (Knight ex Forbes) K. Koch, respectively (M. Suffness, pers. comm.).

From 1960 until 1982, about ½ of the plant samples were supplied to the NCI through a cooperative agreement with the Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA). This agreement, expending nearly ½ million dollars annually since 1972, was terminated as a result of widespread 1981 budget cuts of federal programs. Other substantial suppliers were Commonweath Scientific and Industrial Research Organization (Australia), Central Drug Research Institute (India), National Defense Medical Center (Taiwan), University of Arizona, University of Costa Rica, University of Concepcion (Chile), University of Brazil (Rio de Janeiro), and the University of Hawaii (J. L. Hartwell, pers. comm.).

[1] Received 3 February 1984; accepted 19 January 1985.

[2] Manuscript prepared privately, a contribution of World Botanical Associates, P.O. Box 2829, Laurel, MD 20708-0829. The views expressed and interpretation given are those of the author and do not reflect the official view of the ARS, USDA.

The NCI procedure followed a stepwise-collecting and -testing protocol to isolate methodically bioactive chemicals in plants for ultimate evaluation in the treatment of human cancers. Initially, species were randomly collected in small amounts from ½–2 kg dried weight. Samples of roots, bark, twigs, leaves, or any combination of these were submitted to a routine extraction and prescreening procedure (Suffness and Douros, 1979). Extracts from approximately 10% of 35,000 species tested were active, i.e., these extracts significantly inhibited tumor growth, increased the life span of leukemic mice, and/or were cytotoxic in vitro (Geran et al., 1972). Selected active species were then re-collected in large quantities (50–250 kg dried) to isolate the active agent(s). Occasionally, massive samples of one to many tons were needed to supply sufficient amounts of the active agent for preclinical and/or clinical studies.

In this paper, the terms "tested" and "screened" refer to the "prescreen" procedures (Suffness and Douros, 1982). Abbreviations cited for tumors (screens, bioassays, or test systems) follow Hartwell (1976, Table 2), and include: WA (Walker carcinoma 256, rat), SA (Sarcoma 180, mouse), CA (Adenocarcinoma 755, mouse), KB (Human epidermoid carcinoma of the nasopharynx, cell culture), LE (Lymphoid leukemia L-1210, mouse), LL (Lewis lung carcinoma, mouse), and PS (Lymphocytic leukemia P-388, mouse). The duration for which these were employed in the prescreen is shown in Suffness and Douros (1979, Fig. 2); PS and KB were the tumors primarily used since 1969.

"Random" collecting was not entirely random (Spjut and Perdue, 1976). It is broadly defined in this paper as sampling without a preconceived selection of species. This is not to imply that samples were obtained without thought. An initial tendency was to shortcut the discovery process by collecting plants on the basis of folkloric, chemotaxonomic, and climatic relationships, but, as screening experience progressed, there was a tendency to minimize bias in collecting so that promising compounds would not be missed. Guidelines focused more on excluding rather than selecting plant taxa—species in 1971, families to a limited extent from 1972–1975, and, finally, genera in 1979. Taxa were precluded from further screening for 2 reasons: (1) many samples had been tested without yielding significant activity, or (2) active agents had been isolated and it was apparent that continued screening would not lead to isolation of new compounds. Unless a change was made in the screening methodology, there was little to gain from collecting additional species of figs (*Ficus*), for example, because after testing more than 10% of the species in *Ficus*, it was clearly evident that PS activity was infrequent and unlikely to exceed marginal criteria; similarly in milkweeds (*Asclepias*) activity would predictably occur in KB, and the cytotoxic agents were invariably found to be cardenolides. Genera excluded from further screening are listed in this paper (Table 1, 2).

The random acquisition of higher plant samples was thus carefully guided to avoid duplication of those genera and species already screened. With the exception of excluded plants and those not occurring in sufficient abundance, species were sampled as encountered in selected geographic areas. Additionally, plants in genera not previously tested and others reportedly used for certain medicinal purposes were especially sought out but "random" collecting was not entirely abandoned in lieu of this. These modalities were sometimes combined into an overall strategy

TABLE 1. GENERA WITH 100 OR MORE EXTRACTS SCREENED FOR ANTITUMOR ACTIVITY.[a]

Abies—50
Abutilon—100
Acacia—800
Acalypha—450
Acer—200
Aegiphila—160
Agave—300
Albizia—150
Alchornea—70
Allium—450
Allophylus—190
Alnus—35
Aloe—332
Amaranthus—60
Annona—120
Ardisia—400
Artemisia—400
Asclepias—120
Aspidosperma—80
Aster—500
Astragalus—2,000
Atriplex—200
Baccharis—400
Bauhinia—300
Berberis—450
Betula—60
Bidens—230
Bridelia—60
Buddleja—100
Byrsonima—120
Caesalpinia—100
Calliandra—100
Calyptranthes—100
Canthium—200
Capparis—250
Casearia—160
Cassia—600
Cassine—40
Casuarina—45
Ceanothus—55
Cecropia—100
Celtis—80
Centaurea—600
Cestrum—150
Chenopodium—150
Chrysanthemum—200
Chrysophyllum—150
Cinnamomum—250
Cirsium—150
Cissus—350
Citharexylum—115
Citrus—12
Clematis—250
Clerodendrum—400
Clethra—68
Clusia—145
Coccoloba—150
Combretum—250
Cordia—250
Cornus—4
Crotalaria—650
Croton—750
Cryptocarya—250
Cupania—55
Cyperus—550
Dalbergia—300
Datura—10
Derris—80
Desmodium—450
Dioscorea—600
Diospyros—500
Dombeya—350
Drypetes—200
Elaeocarpus—200
Erigeron—200
Erythrina—100
Erythroxylum—250
Eucalyptus—500
Euclea—20
Eugenia—1,000
Euonymus—176
Eupatorium—1,200
Euphorbia—2,000
Fagara—250
Faramea—120
Ficus—800
Fraxinus—70
Garcinia—400
Gardenia—250
Gnidia—100
Grewia—150
Guarea—170
Gaultheria—250
Helenium—40
Helianthus—110
Helichrysum—500
Heliotropium—250
Hibiscus—300
Hypericum—400
Hyptis—400
Ilex—400
Indigofera—700
Inga—200
Ipomoea—500
Jacaranda—50
Jasminum—300
Jatropha—175
Juniperus—60
Lantana—150
Liatris—40
Linum—230
Litsea—400
Lobelia—300
Lonchocarpus—150
Lonicera—200
Lupinus—200
Macaranga—280
Manilkara—70
Maytenus—225
Miconia—700
Mikania—252
Mimosa—500
Morinda—80
Myrcia—500
Myrica—35
Nectandra—100
Ocotea—400
Oenothera—80
Opuntia—250
Palicourea—200
Passiflora—500
Penstemon—252
Persea—150
Phoradendron—190
Phyllanthus—600
Physalis—100
Phytolacca—35
Pinus—100
Piper—2,000
Pithecellobium—200
Pittosporum—150
Plantago—265
Pluchea—50
Podocarpus—100
Polygonum—300
Populus—35
Potentilla—500
Pouteria—50
Protea—130
Protium—90
Prunus—430
Psidium—140
Psychotria—700
Quercus—450
Randia—300
Rapanea—200
Rhamnus—160
Rhus—250
Rosa—250
Rubus—250
Rumex—200
Salix—500
Salvia—700
Sambucus—40
Sapium—120
Scaevola—100
Senecio—3,000
Sida—200
Siparuna—150
Sloanea—120
Smilax—350
Solanum—1,700
Solidago—100
Sterculia—300
Strychnos—200
Styrax—130
Swartzia—100
Symplocos—350
Syzygium—500
Tabebuia—100
Tabernaemontana—100
Tecoma—16
Tephrosia—300
Terminalia—250
Thalictrum—150
Theobroma—30
Tibouchina—200
Tournefortia—150
Trema—30
Trichilia—300
Vaccinium—400
Verbena—253
Vernonia—1,000
Viburnum—216
Virola—60
Vismia—35
Vitex—250
Vitis—70
Xylopia—150
Yucca—40
Zanthoxylum—30

Total: 201 genera
58,956 species

[a] Numbers following genera are of species in genus (Willis, 1973).

of random collecting that was targeted for geographical areas predetermined to have high concentrations of untested genera and preselected medicinal plants (USDA Memorandum, 1979c).

Retrospective studies on the relationships of antitumor activity with taxonomy

TABLE 2. GENERA CHEMICALLY EVALUATED AND COMPLETED FOR SCREENING IN P388 AND/OR KB.[a]

Acanthospermum—8
Acnistrus—50
Acokanthera—15
Adenium—15
Alangium—17
Allemanda—15
Alstonia—50
Ambrosia—40
Amorpha—20
Androcymbium—35
Apocynum—7
Argemone—10
Arisaema—150
Asarum—70
Austrocedrus—1
Baileya—4
Balduina—3
Baliospermum—6
Begonia—900
Bersama—2
Bleekeria—10
Boehmeria—100
Borreria—150
Bothriospora—1
Brandegea—1
Bryonia—4
Callicarpa—140
Callitris—16
Camptotheca—1
Capirona—5
Catharanthus—5
Cephaelis—180
Chrysothamnus—12
Citrullus—3
Coccinia—31
Colchicum—65
Colubrina—24
Coronilla—20
Cracca—8
Cryptostegia—2
Cucumis—25
Cucurbita—15
Cupressus—20
Cyclea—30
Daphne—70
Daphnopsis—46
Datisca—2
Digitalis—30
Ecballium—1
Echinocystis—15
Edgeworthia—3
Elephantopus—32
Eriope—28
Eriophyllum—11
Ervatamia—80
Excavatia—10
Ferdinandusa—20
Gaillardia—28
Gloriosa—5
Gomphocarpus—50
Haplophyton—3
Hazunta—8
Hedera—15
Hernandia—24
Hillia—20
Holarrhena—20
Hydrastis—2
Hymenoclea—4
Iberis—30
Iphigenia—12
Kanahia—1
Kedrostis—35
Kopsia—25
Kreysigia—1
Lagenaria—6
Lepidium—150
Liriodendron—2
Littonia—8
Lloydia—20
Luffa—6
Mappia—7
Marah—7
Merendera—10
Millettia—180
Momordica—45
Mundulea—31
Muscari—60
Neorautanenia—3
Nerium—3
Nothapodytes—4
Ochrosia—30
Ocimum—150
Ophiorrhiza—150
Ornithoglossum—3
Pachyrhizus—6
Pagiantha—20
Parinari—60
Pedilanthus—14
Peponium—20
Peschiera—25
Phoebe—70
Phormium—2
Piscidia—10
Podanthus—2
Putterlickia—2
Rheedia—45
Rhizophora—7
Rhododendron—600
Richardia—10
Rondeletia—120
Sandersonia—1
Sarracenia—10
Scilla—80
Sicyos—15
Sophora—50
Spatholobus—40
Steganotaenia—2
Stereospermum—24
Strophanthus—60
Taiwania—3
Taxus—10
Teclea—30
Thevetia—9
Thuja—5
Tocoyena—20
Trichosanthes—15
Tricomaria—1
Tylophora—50
Urginea—100
Veratrum—25
Widdringtonia—5
Ziziphus—100

Total: 132 genera
5,460 species

[a] Some genera in Table 1 (e.g., *Maytenus*) also belong here. Numbers following genera are of species in genus (Willis, 1973).

and folklore (Barclay and Perdue, 1976; Spjut and Perdue, 1976) concluded that random collecting is the best approach in a long-term program. Suffness and Douros (1979) also listed advantages and disadvantages for testing at random (random screen) as compared to selective screening; they indicated random screening is preferable because of low procurement costs due to "plants readily available" and because of a possibility of chance discovery of novel anticancer drugs. The number of plants readily available was not then known. This has been difficult to determine because there is considerable nomenclatural synonymy among approximately ¾ million species of flowering plants that have been described, while estimates for the actual number of angiosperm species varies from ⅕ million to more than ½ million. The number of species available to the NCI program will be assessed here in terms of Good's (1974, also citing Lemée, 1929–1943, and Willis, 1966) estimate of 225,000 flowering plant species. Good's estimate is included in Table 3, which shows the numbers of genera and species available and screened for each division of higher plants. Of a total of about 235,000 existing species, only 15% have been screened.

TABLE 3. GENERA/SPECIES OF HIGHER PLANTS SCREENED FOR ANTITUMOR ACTIVITY.

| Division | Total available[a] | Screened[b] |
|---|---|---|
| Psilopsids (Psilophyta) | 2/10 | 1/1 |
| Lycopods (Lycopodophyta) | 5/950 | 3/53 |
| Horsetails (Arthrophyta) | 1/23 | 1/12 |
| Ferns (Pterophyta) | 225/8,500 | 107/315 |
| Cycads (Cycadophyta) | 10/100 | 9/37 |
| Gingos (Gingophyta) | 1/1 | 1/1 |
| Conifers (Coniferophyta) | 50/400 | 49/369 |
| Gnetophytes (Gnetophyta) | 3/73 | 2/25 |
| Flowering plants (Magnoliophyta) | 12,500/225,000 | 5,500/35,000 |

[a] Estimates from Willis (1973), Holttum (1973), Scagel et al. (1965), Bold (1967), and Good (1974).

[b] Recorded from the NCI Plant Header File (1982) except for flowering plants. Species estimate of flowering plants from Suffness and Douros (1982). Genera of flowering plants extrapolated from a sampling of 90 of 179 pages of the NCI file (1978), synonymy eliminated, and additional genera recorded from USDA accession records. The USDA collections have vouchers deposited at the National Arboretum (NA), Washington, DC, but these represent about ½ of the NCI samples extracted.

With 85% of the world flora of higher plants still untested (Table 3), it may appear premature for the NCI and the USDA to impose guidelines to avoid duplication in collecting. However, there are limitations to what is practically collectable. In the sections that follow, these limitations and guidelines will be reviewed in relationship to antitumor activity and plant classification, phytogeographic vs. political boundaries, geographical areas explored, and distribution of common genera and species. As a result, I will suggest a feasible number of species collectable for the NCI random-acquisition program, show that a disproportionately large number of species are distributed in relatively few genera, demonstrate the impact of guidelines on procurement, and illustrate the importance of phytogeography in selecting areas for plant exploration and screening. Finally, a synthesis of these points should demonstrate the potential utility of a proposed manual to identify genera and species commonly encountered in tropical and temperate regions of the world.

## ANTITUMOR ACTIVITY AS APPLIED TO PLANT CLASSIFICATION

Hartwell (1976) and Suffness and Douros (1979) portray a diversity of compounds having shown antitumor activity and/or cytoxicity. This diversity is also the product of an evolving screen. Early in the program, active agents included tannins, phytosterols, and other compounds that were later found to have no chemotherapeutic potential. As a result, it was necessary to modify the extraction procedure and/or employ new tumor systems to increase the chances of discovering useful anticancer agents. From such changes, it is important to recognize that species with negative test results under one methodology may later show significant activity.

Biological activity also varies with plant parts (e.g., root, stembark, twig, etc.) and conditions under which samples are obtained, such as the location or soil type and developmental stage of the plant (Croom, 1983). For instance, KB activity was reported in only one of 2 stem samples of white-stemmed milkweed (*Asclepias albicans* S. Wats.) collected from the same location (near Desert Center, CA) but

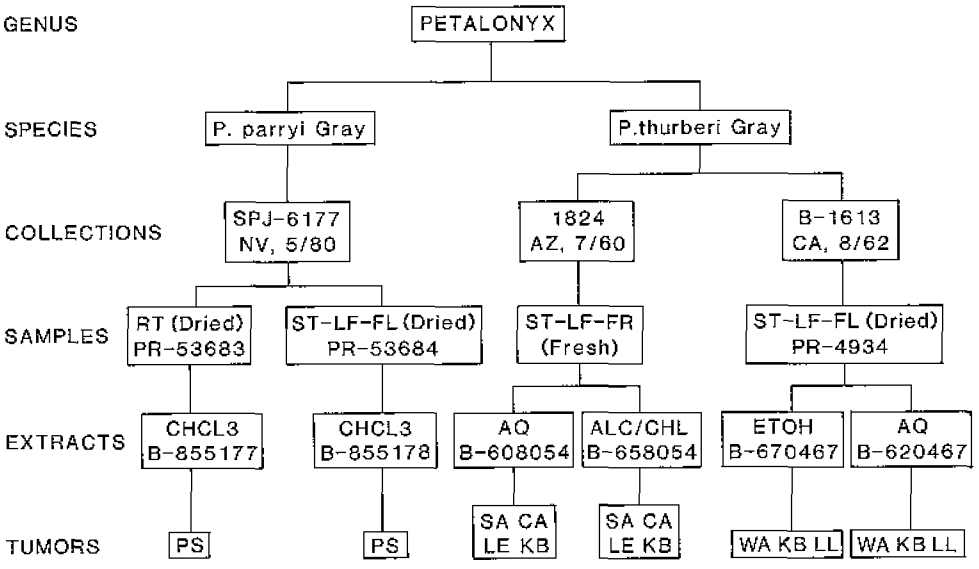


**Fig. 1.** Variables in the NCI antitumor screening of higher plants. Data in boxes are from the NCI Plant Header File and the NCI Screening Data Summary.

COLLECTIONS: SPJ = R. W. Spjut (Collector); B = A. S. Barclay; NV = Nevada; AZ = Arizona; CA = California. SAMPLES: RT = root; ST-LF-FL = stem-leaf-flower (combined); FR = fruit; PR = standard prefix for samples accessioned by the USDA. EXTRACTS: B = standard prefix for the NCI crude extracts prepared from natural products; $CHCl_3$ = chloroform; EtOH = ethanol; AQ = aqueous; ALC/CHl = ethanol/chloroform. TUMORS: PS = Lymphocytic leukemia P-388 (mouse); SA = Sarcoma 180 (mouse); CA = Adenocarcinoma 755 (mouse); LE = Lymphoid leukemia L-1210 (mouse); KB = Human epidermoid carcinoma of the nasopharynx (cell culture); WA = Walker carcinoma 256 (rat); LL = Lewis lung carcinoma (mouse).

at different times (Feb. 12 and March 28, 1972; USDA Memorandum, 1976). The active sample (*Spjut 2189,* NA, HSC) had flowers in bud; the inactive one (*Spjut 2239,* NA, HSC) had fruits. In another milkweed (*Asclepias eriocarpa* Benth.), Nelson et al. (1981) found that the cardenolide content changed significantly during the growing season. Cassady and Suffness (1980) have summarized the stereochemical requirements for "cardiotonic activity" and concluded, for example, a large loss in activity occurs when "a 14,15 unsaturation is introduced" in the basic molecule.

Fig. 1 illustrates a hierarchial relationship of the variables from genus to the tumor level, and an expression of this variability, as applied to medicinal plants, is presented in Table 4, for which Watt and Breyer-Brandwijk (1962) was reviewed for medicinal or poisonous uses in 443 species randomly collected in 1971 from the Southern Highlands of Tanzania. The review was undertaken for Spjut and Perdue (1976) but data were never published. Of a total of 143 medicinal species recorded, 119 were found to have been collected on prior occasions: from Tanzania, Kenya, Uganda, Ethiopia, India, or some other tropical source illustrated in Perdue (1976, Fig. 8). These medicinal species were grouped as follows: 24 collected for the first time, 29 collected twice, 24 collected 3 or 4 times, 32 collected 5 or 6 times, 17 collected 7 or 8 times, and 17 collected 9 or more times. The percentage of active species was determined for each class (group). Data (Table

TABLE 4. ANTITUMOR ACTIVITY IN SPECIES OF MEDICINAL AND POISONOUS PLANTS ACCORDING TO THE NUMBERS OF COLLECTIONS.

| Number of collections | Number of species | Active species[a] | |
|---|---|---|---|
| | | All tumors | PS & KB |
| 1 | 24 | 3 (12.5%) | 3 (12.5%) |
| 2 | 29 | 4 (13.8%) | 4 (13.8%) |
| 3–4 | 24 | 13 (54.2%) | 7 (29.2%) |
| 5–6 | 32 | 12 (37.5%) | 9 (28.1%) |
| 7–8 | 17 | 8 (47.1%) | 4 (23.5%) |
| 9+ | 17 | 4 (23.5%) | 1 (5.9%) |

[a] PS = P388 Lymphocytic leukemia (mouse); KB = Human epidermoid carcinoma of the nasopharynx (cell culture); see also Suffness & Douros, 1979 (p. 84, Fig. 2) for other tumors and duration employed. Percentages of active species are cumulative. For example, 29.2% of the species collected 3 or 4 times were active in PS and/or KB, but if based on the total number of collections (≈24 × 3.5), this is near 8.3%.

4) suggest that a species should not be collected for the NCI antitumor prescreen more than 7 or 8 times, a point of diminishing returns for discovering activity in a species.

The systematic distribution of active agents may correlate with morphological characteristics that distinquish taxonomic levels of higher plants (division, class, order, family, genus, and species). Cucurbitacins have been isolated from many genera but most of these occur in the Cucurbitales (of Hutchinson, 1959). Activity due to quassinoids has been restricted to the Simaroubaceae, where this activity has occurred in more than 25% of the species tested. Cytotoxic lignans were frequently found in conifers, within the Cupressaceae, but occasionally in flowering plants, notably the Burseraceae. While such relationships are occasionally evident at the family level or above, these are more often clearer at the genus level. If 2 or more species of a genus are active (based on the same screening procedure), identical or similar compounds are likely to be isolated. But, if two or more genera of the same family, or higher taxon, are active, the active chemical structures are likely to be quite different. The genus, therefore, is suggested as the lowest taxon that correlates with chemical diversity.

Data in Table 5 exemplify a correlation of antitumor activity with genera as related to their size. Data are from compilations of families of flowering plants that were utilized by Barclay and Perdue (1976), but are here rearranged. The distribution of 3,533 genera is shown according to the number of species tested in each genus for 15,589 species of flowering plants. Active species are those from the NCI cumulative list (1974). Probable (or expected) percentages of active genera were determined from the percentages of active species for each class (number of species tested per genus). In virtually all classes, actual percentages of active genera are less than the probable, suggesting that activity in genera is not at random. Also, monotypes (genera with only one species tested), with 5 exceptions, have the lowest percentage of active species, and genera with 2–7 species tested have the highest percentages of active species. This indicates a higher than average chance of finding activity in more than one species of an active genus, but also the chance of finding activity in a genus diminishes if 7 species have been screened without success.

In summary, antitumor activity best correlates with phytochemical diversity at

TABLE 5. PROBABLE AND ACTUAL PERCENTAGES OF ACTIVE GENERA ACCORDING TO NUMBER OF SPECIES TESTED IN A GENUS.

| Number of species tested per genus (class) | Number of genera | % of active species | Probable % of active genera | Actual % of active genera |
|---|---|---|---|---|
| 1 | 1,672 | 6.8 | 6.8 | 6.8 |
| 2 | 602 | 9.5 | 18.1 | 17.4 |
| 3 | 289 | 9.1 | 24.9 | 24.9 |
| 4 | 185 | 10.9 | 37.0 | 28.1 |
| 5 | 146 | 7.4 | 31.9 | 25.3 |
| 6 | 85 | 10.2 | 47.6 | 38.3 |
| 7 | 88 | 11.2 | 56.5 | 44.3 |
| 8 | 56 | 6.9 | 43.6 | 39.3 |
| 9 | 54 | 7.8 | 51.9 | 38.9 |
| 10 | 30 | 9.0 | 61.1 | 46.7 |
| 11 | 39 | 7.9 | 59.6 | 56.4 |
| 12 | 23 | 7.2 | 59.2 | 47.8 |
| 13 | 29 | 7.7 | 64.7 | 51.7 |
| 14 | 17 | 4.2 | 45.2 | 41.2 |
| 15 | 16 | 5.8 | 59.2 | 43.8 |
| 16 | 22 | 7.7 | 72.3 | 63.6 |
| 17 | 10 | 7.1 | 71.4 | 70.0 |
| 18 | 17 | 6.9 | 72.4 | 61.1 |
| 19 | 15 | 6.3 | 71.0 | 63.2 |
| 20–22 | 29 | 8.4 | 84.2 | 75.9 |
| 23–25 | 20 | 9.1 | 89.9 | 70.0 |
| 26–29 | 28 | 5.8 | 80.7 | 67.9 |
| 30–39 | 22 | 9.7 | 97.0 | 86.4 |
| 40–59 | 20 | 6.3 | 96.0 | 95.0 |
| 60–99 | 14 | 7.2 | 99.7 | 100.0 |
| 100+ | 5 | 9.4 | ~100.0 | 100.0 |

*Columns 1 and 2* illustrate Willis' (1922) hollow curve distribution. For example, there were 1,672 genera with only a single species tested (=1,672 species) and 5 genera with at least 100 species tested (=500 + species). In essence, the bulk of the genera screened were rarely collected, but species most frequently collected belong to relatively few genera.

*Column 3.* Activity includes any tumor system used in the NCI screen. Active agents represent nearly all classes of chemical compounds (Hartwell, 1976). Percentages of active species are from 2–3% less than in Barclay and Perdue (1976) because they updated their findings with new reports of active species until presentation of their paper (June, 1975).

*Column 4.* Probable percentages of active genera were determined from percentages of active species found in each class (column 3) using the binomial expansion of $(p + q)$ where n = number of species tested; p = frequency of activity in species; and q = frequency at which activity does not occur in species. The 6.8% activity for monotypic genera might approximate the real frequency of biological activity in natural products; however, the actual percent for each class was used to represent 'p' instead of applying 6.8% as a constant for all classes.

*Column 5.* Actual percentages of active genera in many classes, especially in the middle range, are less than the probable, indicating that activity is not entirely at random.

the genus level, but this is not always clear-cut because (1) the same, or closely related, antitumor compounds are sometimes distributed among many genera of a family (e.g., quassinoids) or higher taxon, (2) bioassays have been sensitive to ubiquitous classes of chemicals (e.g, tannins), and (3) changes in screening methodology have resulted in discovery of novel compounds that do not repeat the taxonomic distribution pattern seen for other active agents isolated from the same or closely related taxa. The latter is exemplified by active proteins almost exclusively discovered by Dr. Jack Cole, University of Arizona (Tables 3 and 18 of Hartwell, 1976). This discovery appears to be a product of the extraction procedure (Fig. 2, Statz and Coon, 1976) and bioassays employed (SA, CA, LL, WA, and

KB). Much of the activity in the Asteraceae has been attributed to sesquiterpene lactones, the major exception being the pyrrolizidine alkaloids in the tribe Senecioneae, but a change in methodology could discover novel chemical structures characteristic of one to many species of a genus or tribe. Alkaloids, diterpenes, and possibly other classes of chemotherapeutic interest, appear widespread in the Asteraceae and show chemotaxonomic differentiation at the tribe, genus or species level (Hegnauer, 1977).

## PHYTOGEOGRAPHICAL VS. POLITICAL BOUNDARIES

Floras are usually defined by political boundaries that bear little relationship to the natural distribution of the plants. This is often necessary since the natural boundaries cannot be recognized until taxa are described and their geographic distributions plotted. Vast areas, especially in South America, are still poorly known, floristically (Gentry, 1978).

To collect the greatest diversity for a random-screening program, a phytogeographic approach should be employed. A widely known classification is that of R. Good (1964), who divided the world into 8 floristic kingdoms or subkingdoms, 37 floristic regions, and 130 provinces. A more recent edition of Good (1974) shows no change with respect to his classification and map. It should be noted that Good equated small areas such as the Cape Region of South Africa (South African Kingdom) with temperate Europe, Asia, and North America combined (Boreal Kingdom). The South African flora is remarkable for its endemism and richness, where 29% of 1,930 genera and 80% of 18,500 species are endemic (Goldblatt, 1978).

Fig. 2, prepared in December 1982, shows 58 floristic regions recognized for a genera-based methodology screen described in Suffness and Douros (1982). Floristic regions reflect a modification of Good (1964) by the ecological approach of White (1970) and were drawn from knowledge and consultation with many references, particularly Champion and Seth (1968), Mani (1974), Zohary (1973), Van Steenis (1948–1949), Burbidge (1960), Cochrane (1963), Brenan (1978), Quézel (1978), L'Association pour L'Etude Taxonomique de la Flora d'Afrique Tropicale (1959), Hueck (1966), Hueck and Seibert (1972), and Prance (1977). The depicted zones represent a generalized summary of collections made for the NCI program. I will later describe a rationale for this in relation to the distribution of the most commonly collected genera and species and untested genera.

Floristic classifications (e.g., Good, 1964; Takhtajan, 1969; Dasmann, 1973; Udvardy, 1975) vary according to purpose and interpretation of the phytogeographic units. For example, Good (1964) did not demarcate North American xerophilous floras (No. 10, 11 in Fig. 2), while in Takhtajan (1969) these were not only recognized but elevated to subkingdom level (Madrean). In Good (1964), the desert floras are part of his Pacific North American Region (No. 5, 10, 11 in Fig. 2), which I refer to as western North America in the following section (Table 6). Data compiled from floristic literature as defined by political boundaries (Rosaceae in Table 6) are more adaptable to this phytogeographic division. The boreal American element is also present in mountainous areas of western North America, but this is not shown in Fig. 2 due to the scale of the map and greater dominance by the other floristic elements.

For practical and political reasons, political boundaries are the operational bases for planning and conducting plant explorations. By applying knowledge of phytogeography, field work can be targeted to selected countries to acquire the greatest diversity with minimum cost and duplication. An example is the country of South Africa that includes the outstanding Cape flora and elements of most of the floristic regions in continental Africa (Fig. 2), or the state of California where 4 of 6 major phytochoria of North America are well represented.

## COLLECTABLE VS. RARE SPECIES

The chief economic limitation to a random collecting methodology is its diminishing returns; only a small fraction of the plants in a defined area is available at a reasonable cost. The uncollectable species are either (1) rare, or (2) if common, then are not feasible to gather in quantity. For example, ladies slippers (*Cypripedium* spp.) are usually rare. An English daisy (*Bellis perennis* L.) is common in lawns and pastures, but it has not been collected because apparently it is too small to be easily gathered in quantity. Generally, a collectable species is one in which 1 man-hr of collecting will yield a minimum of 1 kg for any combination of dried plant parts.

The collectable part of a flora is estimated below on the basis of geographic areas judged extensively collected for the NCI, largely on the author's cumulative years of plant exploration in Africa, Australia, Mexico, and the United States. An entire flora and families of other floras, as published in the literature, were compared against the NCI record (1982) to determine which species have been collected. The results are shown in Table 6.

The number of extracts from samples obtained within each area is also shown (Table 6). Numbers of species collected from specific countries or states are retrievable, but due to synonymy and many samples from arboreta, nurseries, and USDA Plant Introduction Stations, this datum would not be truly representative of the indigenous flora. Numbers of extracts tested, however, are useful for making relative comparisons on collecting in geographical areas. For an estimate of the species represented, I usually divide the number of extracts tested by 3.

Data suggest that it is possible to collect 50% of a flora; however, the percentage will likely vary according to the degree of endemism. Willis (1915, 1916) has shown that endemics are generally rare, in contrast with his common species. Also, Willis' common or widespread species seem to compare favorably with collectable species. For example, Table 2 of Willis (1915) has 53% of the species in a Sri Lanka (Ceylon) flora as "very common," "common," and "rather common," leaving 47% as "rather rare," "rare" and "very rare." The percentage of common Sri Lanka species (53%) compares closely to collectable species for California (51%), possibly because both areas are about equal in percentage of endemic species—30% (Willis, 1915; Raven, 1977). Where endemic species are concentrated in small areas, as in the Cape Region of South Africa (80%), percentages of collectable species will be less.

In Table 6 are Rosaceae species from California (3,625 extracts tested) as compared to all of western North America, north of Mexico (11,546 extracts tested). Of 149 species in *A California Flora* (Munz, 1959), 76 (51%) were screened but only 40 (27%) were actually obtained within California. Obviously, this difference

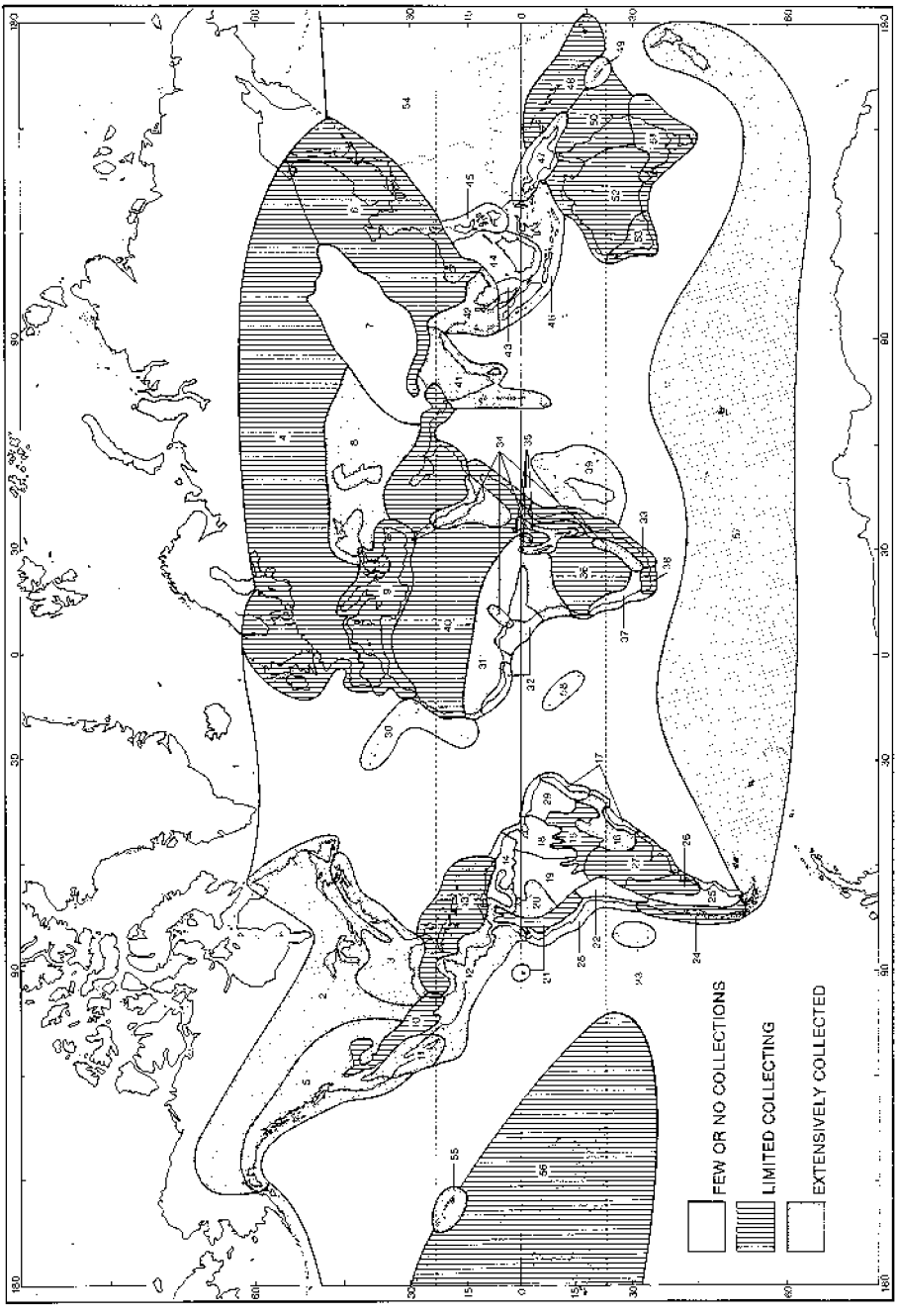
FEW OR NO COLLECTIONS
LIMITED COLLECTING
EXTENSIVELY COLLECTED

TABLE 6. AREAS COLLECTED EXTENSIVELY FOR NCI ANTITUMOR SCREENING.

| Area and samples (extracts) collected | Species recorded from reference | Species collected | |
|---|---|---|---|
| | | Within area | From any location |
| California—3,625, Rosaceae[a] | 149 | 40 (27%) | 76 (51%) |
| Western North America (North of Mexico)—11,546, Rosaceae[b] | 265 | 104 (39%) | 116 (44%) |
| Kenya (uplands)—5,015, Acanthaceae[c] | 123 | 34 (28%) | 42 (34%) |
| India—7,085, Asclepiadaceae and Acanthaceae (Hassan District)[d] | 81 | — | 31 (38%) |
| Panama—2,740, Verbenaceae and Asclepiadaceae[e] | 111 | 17 (15%) | 53 (48%) |
| Sonoran Desert[f]—4,000 | 2,883 | — | 1,319 (46%) |

[a] Munz (1959); [b] MAB/FNA, Council (1978)—a similar count was also obtained from reviewing 10 other localized floras; [c] Agnew (1974), Dale and Greenway (1961); [d] Ramamoorthy (1976), Stevens (1976); [e] Moldenke (1973), Spellman (1975); [f] Shreve and Wiggins (1964). Number of extracts tested for Sonoran Desert is an estimate.

is partly due to Californian species that are more common outside California; however, in extending the area of Rosaceae to other states, there are an additional 116 species of which only 40 (34%) were collected. Two-thirds of the additional species are apparently rare endemics.

The preceding example contrasts collectable vs. rare species much as Willis (1922) compares "rare" vs. "common," or "endemics" vs. "wides." The extended area (California to western North America) does not overlap into other floristic

←

**Fig. 2.** A phytogeographic summary of collections for the NCI Anticancer Screening Program.

Key to floristic regions

1. Arctic & Subarctic
2. Boreal America
3. Southeastern U.S.
4. Boreal Eurasia
5. Pacific Northwest
6. Sino-Japanese
7. Central Asia
8. Southwestern Asia
9. Mediterranean
10. Chihuahuan & Great Basin Deserts
11. Californian Deserts & Chaparral
12. Central America
13. West Indies
14. Venezuela & Guyana
15. Brazilian Savanna
16. Brazilian Highland Forests
17. Brazilian Coastal Rainforests
18. East Amazon
19. Central Amazon
20. West Amazon
21. North Andes & Galapagos
22. Puna
23. Juan Fernandez
24. South Andes
25. Patagonia
26. Monte
27. Chacos & Pampas
28. Pacific Deserts
29. Caatinga
30. Macronesian
31. West African Forests & Savanna
32. Equatorial African Rain Forests
33. Usambara-Zululand
34. Afro-Montane
35. Afro-Alpine
36. Zambezi
37. Namibian
38. African Cape
39. Madagascar
40. African-Indian Desert
41. India & Ceylon
42. Indochina
43. Malaya
44. Borneo
45. Philippines
46. Sumatra, Java, Celebes & Moluccas
47. New Guinea
48. Solomon & Fiji
49. New Caledonia
50. Northeast Australia
51. Southeast Australia & Tasmania
52. Central Australia
53. Western Australia
54. Melanesia & Micronesia
55. Hawaii
56. Polynesia
57. New Zealand & South Temperate Islands
58. Ascension & St. Helena

regions; it encompasses only those regions already explored (No. 5, 10, 11 in Fig. 2). Most additional Rosaceae species are expectedly rare since the widespread ones were collected. As new floristic boundaries are crossed, many new common species are encountered.

The large difference between the number of species collected in Panama (17) and outside Panama (53 minus 17) is attributed to many samples being obtained from Costa Rica (6,870 extracts tested), which belongs to the same floristic region as Panama (Central America, #12 in Fig. 2).

The Sonoran Desert is represented by collections from Mexico (Sonora, Sinaloa, Baja California—905 extracts tested) and the United States (Arizona and California—about 3,100 extracts tested). Applying Shreve and Wiggins (1964), 46% of 2,883 species cited (excluding varieties and subspecies) were tested. Shreve (1964) also enumerates 372 species that make up the vegetation in the Sonoran Desert; of these, 77% were screened. Thus, most of the characteristic species were sampled, and, of those that were not (54%), many are undoubtedly rare.

In addition to rare species, some common grasses (Poaceae), annual herbs, small succulents, and epiphytes are difficult to gather in quantity. The bulk of these occur in 6 families. Table 7 provides a world estimate of major families least likely to be collectable and the extent these were once collected. Barclay and Perdue (1976) reported that 10% of the flowering plants were tested, but restricting their data to families in Table 7, only 2.5% of an estimated 35,600 available species were screened. Extrapolating from a 50% collectable yield for other families (Table 6), 31,850 species in 6 major families were not likely to be obtained for the NCI [$2.5\%/x = 10\%/50\%$; $x = 12.5\%$; $35{,}600 - (12.5\% \times 35{,}600) = 31{,}150$].

Perhaps 100,000 species of higher plants are feasible for collection ($235{,}000 - 31{,}150 \times 50\%$).

## COMMON SPECIES AND GENERA

### *Criteria for common species*

The NCI first limited duplication of screening of species to 10 extracts (NIH-HEW Memorandum, 1971) based on a review of a distribution of all active species according to numbers of extracts tested (Hartwell, pers. comm.).

Determining the number of times a species has been collected is arduous since the information is currently unretrievable by computer. The NCI keyed their data to extract ("B") numbers from which computer counts were made. Rejection of species based on extract counts allowed herbaceous species to be collected more often than woody plants. For example, Californian redwood (*Sequoia sempervirens* [Lamb. ex D. Don] Endl.) could be separated into 4 samples (root, bark, twig, leaf), and, if sampled similarly on 3 occasions, a cumulative of 12 extracts might be listed in the NCI record. Samples of a dandelion (*Taraxacum officinale* Weber) are most likely to consist of the entire plant; thus, it could be collected 10 times before rejection from screening.

In 1978, the allowable number of extracts tested per species was reduced from 10 to 6. This had little additional effect on rejecting species because prior to 1966 two extracts were usually prepared from each sample. Species, on the average, were collected 2 or 3 times before rejection. This guideline is generally supported by data in Table 4.

TABLE 7. FAMILIES LEAST LIKELY TO BE COLLECTED FOR NCI SCREENING.

| Family | Genera/species | |
|---|---|---|
| | Total available[a] | Screened[b] |
| Araceae | 115/2,000 | 38/101 (5%) |
| Aizoaceae | 130/1,200 | 14/37 (3.1%) |
| Bromeliaceae | 44/1,400 | 15/45 (3.2%) |
| Cyperaceae | 90/4,000 | 35/158 (4%) |
| Orchidaceae | 735/17,000 | 48/77 (<½%) |
| Poaceae | 620/10,000 | 150/500 (5%) |
| Total | 1,734/35,600 | 300/918 (2.5%) |

[a] From Willis (1973).

[b] Data compiled in 1974. Barclay and Perdue (1976) reported about 10% of the flowering plants were screened for antitumor activity. Percentages are of species screened.

An NCI computer list, dated April 8, 1981, included about 6,200 species with 6 or more extracts tested. These, then, are the common species.

*Criteria for common genera*

A generic approach to the collecting of plant samples was initiated in May 1978 (USDA Memorandum, 1978a). By January (NCI Memorandum, 1979a), 201 genera were precluded from further screening on the basis of 100 or more extracts tested, which is equivalent to 16 or more species tested in Table 5. The 201 genera are listed in Table 1 with an estimated number of species for each genus (Willis, 1973). Collectively, these genera account for 58,956 species, or ¼ of the world flora of higher plants. Additionally, the NCI considered another 132 genera as having completed screening based on tumor activity and subsequent isolation of active compounds (Table 2).

Willis (1922) had reported that the majority of 12,571 genera are monotypes (4,853 or 38.6% with one species per genus) or ditypes (1,632 or 12.9%). From numbers of species tested per genus (Table 5), 47% of all genera screened were monotypes, but most species collected occur in fewer than 330 genera (11 or more species tested). One-third of all species screened belong to the 201 genera in Table 1 (NCI Memorandum, 1979b).

*Impact of common species and genera: a phytogeographic summary*

Genera and species excluded from the NCI screen were combined (USDA memorandum, 1979a) into a single listing known as "SLOP," an acronym for "Species Low On Priority" (Suffness and Douros, 1982). A USDA list (1980b) precluded further testing of all species in 333 genera (Table 1, 2) and another 2,905 species in 1,773 other genera. This (SLOP) included most weeds (e.g, *Achillea millefolium* L., *Agropyron repens* (L.) Beauv., *Agrostemma githago* L.; 77% of 224 in a USDA handbook of selected weeds, 1970), most species of major economic importance (e.g, *Beta vulgaris* L., *Cannabis sativa* L., *Capsicum frutescens* L., *Carthamus tinctorius* L.; 53% of Terrell's checklist of 3,000 economic plants, 1977), and many naturalized or commonly cultivated shrubs or trees (e.g., *Ailanthus altissima* (Mill.) Swingle, *Azadirachta indica* A. Juss.), and other wide-

spread tropical species (e.g., *Apodytes dimidiata* E. Mey. ex Arn., *Bixa orellana* L., *Crescentia cujete* L., *Curatella americana* L., *Hunteria zeylanica* (Retz.) Gard. ex. Thw., *Parinari excelsa* Sabine, *Scutia myrtina* (N.L. Burm. f.) Kurz, *Urera baccifera* (L.) Gaud.; 71% of 115 pantropical species in Good, 1974).

Table 8 shows the results of a comparison of (1) low-priority taxa (SLOP), and (2) genera collected for the NCI with (3) the available literature on vegetation for selected areas of the world. This provides an example as applied to a generalized phytogeographic summary of exploration for the NCI depicted in Fig. 2. The stippled zone represents extensive collections, the striped indicates limited collecting, and clear zone yields the highest percentages of new genera with a minimum of low-priority taxa (SLOP).

Nicaragua and Malawi (Table 8) are examples of countries where major collections for the NCI have not been made, but occur in an extensively collected zone because 70% of random collections are expected to be low priority taxa (SLOP). This illustrates the importance of phytogeography in selecting areas for plant exploration.

Percentages of SLOP would be higher than indicated for other countries, such as Cameroun and Sudan in the Afromontane zone, if data were limited to woody plants. As indicated earlier, a disadvantage with extracts (or samples) as a data base is that herbaceous species can be repeatedly collected more than woody plants before being classified as low priority.

On the other hand, the distribution of low priority taxa (SLOP) is correlated with dominance. For example, 35% of the species in an Ecuadorian flora (Rio-Palenque, Dodson and Gentry, 1978) were identified as SLOP, but according to the composition of its rain forest vegetation, SLOP were found in 70% of the canopy trees, 51% and 31% of the understory trees and shrubs, respectively, and only 7% of the herbs.

Thus, in regions where collecting has been extensive, low priority taxa characterize much of the vegetation. In the chaparral and coastal sage vegetation of California, this includes species in small genera: *Adenostoma fasciculatum* Hook. & Arn., *Garrya flavescens* S. Wats., *Dendromecon rigida* Benth., *Eriophyllum confertiflorum* (DC.) A. Gray, *Eriodictyon californicum* (Hook. & Arn.) Torr., *Heteromeles arbutifolia* (Dryander in W. Ait. f.) M. Roem.; those in larger genera: *Arctostaphylos patula* Greene, *A. uva-ursi* (L.) Spreng, *A. drupacea* (Parry) Macbr. and *A. pungens* H.B.K.; and all species in *Ceanothus, Rhamnus, Rhus, Salvia, Baccharis, Quercus,* and *Pinus.* Species of a more localized occurrence are often not included, such as *Adenostoma sparsiflorum* Torr. This is a tree-like shrub that extends from Los Angeles to 200 mi south of San Diego (in Baja California) and occasionally forms pure stands, as in the San Jacinto Mountains; nevertheless, it has been screened. The one other species of *Adenostoma,* already mentioned as low priority, is quite common from northern Baja California to near Redding in northern California.

## DISCUSSION AND CONCLUSIONS

Selecting plants at random for screening of their chemicals is possibly the most efficient approach to discover new biologically active compounds, provided that the limitations are taken into consideration. I have suggested 100,000 species of

TABLE 8. EXAMPLES OF REGIONS CORRESPONDING TO ZONES IN FIG. 2.

| Zone | No. of samples (extracts) collected from each country or state | Total species recorded from reference cited | New genera[a] | SLOP[b] |
|---|---|---|---|---|
| Extensively collected | | | | |
| Central America | | | | |
| Nicaragua (Taylor, 1963) | 25 | 195 | 3% | 72% |
| Afro-montane | | | | |
| Malawi (Chapman and White, 1970) | 0 | 180 | 4% | 69% |
| Cameroun (Richards, 1963) | 17 | 85 | 4% | 49% |
| Sudan (Jackson, 1956) | 29 | 149 | 8% | 40% |
| Partially collected | | | | |
| Zambezi | | | | |
| Zaire (Mullenders, 1954) | 4 | 175 | 10% | 35% |
| Monte | | | | |
| Argentina (Morello, 1958) | 26 | 308 | 10% | 25% |
| Mediterranean | | | | |
| Morocco (Emberger, 1939) | 2 | 154 | 10% | 24% |
| Southeast Australia and Tasmania | | | | |
| New South Wales (Fraser and Vickery, 1938) | 396 | 169 | 6% | 19% |
| Little collected | | | | |
| Equatorial African Rain Forest | | | | |
| Cameroun (Letouzey, 1968) | 17 | 147 | 22% | 35% |
| Ivory Coast (Mangenot, 1955) | 24 | 155 | 14% | 31% |
| New Caledonia | | | | |
| New Caledonia (Jaffré and Latham, 1974) | 150 | 90 | 16% | 23% |
| Venezuela and Guyana | | | | |
| Guyana (Davis and Richards, 1934) | 89 | 122 | 15% | 34% |
| Borneo | | | | |
| Borneo (Bruning, 1965) | 39 | 160 | 13% | 24% |
| Malaya | | | | |
| Malaya (Poore, 1968) | 113 | 232 | 7% | 19% |
| Western Australia | | | | |
| Western Australia (Beard, 1976) | 289 | 134 | 9% | 12% |

[a] "New Genera" are those not found in a list (NCI, 1980) of all plant genera tested for antitumor activity. % = # new genera/# species recorded.

[b] "SLOP" = Species Low on Priority, a composite listing of 333 genera and 2,905 species in 1,773 genera (USDA, 1980b).

higher plants are practically collectable for the NCI anticancer screening program. To avoid duplication and to increase cost effectiveness in the NCI screening, guidelines evolved to preclude further screening of 3,238 low priority taxa (all species in 333 genera, and 2,905 species in 1,773 genera). Because ⅓ of all species tested belong to 201 common genera (Table 1), the collectable number of species might be adjusted as follows:

1. Total number of species available (Table 3) is 235,000. Subtract species in common genera (58,956 in Table 1), and species in families least likely to be collected (35,600 in Table 7) = 140,444.
2. Total number of collectable species is 140,444 × 50% = 70,222, except for species in Tables 7 and 1.

3. One third of all species collected belong to genera in Table 1; thus, ⅓ of 35,000 screened (Table 3) = 11,667.
4. Total collectable in those families least likely to be collected (Table 7) is 35,600 − 31,150 = 4,450.
5. Adjusted total of collectable species is the sum of 2, 3, and 4 above, or 86,339.

Approximately 85,000 species might then be viewed as a long term goal of a random-screening program.

The impact of low priority taxa (SLOP) guidelines on plant procurement varied geographically, as evident in Table 8. The author obtained samples randomly from the southwestern United States at $15 each (USDA Memorandum, 1979b), but in having to avoid low priority taxa similar samples cost $30 each (USDA Travel Report, 1981a; USDA Memorandum, 1979b). Western Australia (Table 8) has the lowest percentage of "SLOP" (12%) and the author's collections from there cost $18 each (USDA Travel Report, 1981b). These costs exclude overhead, salary, and shipping. Further exploration is not recommended in extensively collected areas unless projects are primarily undertaken for re-collections or massive samples. Over-collecting of genera and species has to be avoided in view of high screening and processing costs—$300 per extract (Suffness, pers. comm.).

Willis (1916, 1922) frequently published tables showing numbers of species distributed according to various sizes of geographic areas. He first sorted species into 2 categories: "endemics" (restricted to the area of study) and "wides" (occurring within and outside the area of study). With endemic species, the largest proportion were always in his rare classes, or of the narrower ranges in geographic distribution; but with the "wides," most were "very common," or in the classes of wide geographic distribution. Except for the clear area in Fig. 2, the common species were screened; the majority of untested genera are probably monotypic and rare.

Not usually recognized as a category of widespread taxa are compilations of medicinal plants. In the Philippines flora, more than 75% of the species are endemic (Good, 1974) yet only 7% are in Quisumbing's (1951) medicinal compendium of the Philippines. Moreover, 73% of the 855 Quisumbing plants were screened from samples primarily collected outside of the Philippines (unpublished data for Spjut and Perdue, 1976). A survey of more than 3,000 species used against cancer (Hartwell, 1967–1971) includes 450 genera that are known to occur in North America but only 5 of these were never collected, as compared to more than 150 untested genera compiled from floristic literature of the western United States (USDA Memoranda, 1978b,c, 1980a). Similarly, the series of Tanzanian collections employed for data in Table 4 includes 143 medicinal species that are the "wides" and 300 not reported in Watt and Breyer-Brandwijk (1962) that correspond to Willis' "endemics." Most species reportedly used in medicinal folklore are widely distributed; therefore, these have been screened for antitumor activity as a consequence of a random acquisition program.

Three interrelated distribution types impose limitations to a random screening methodology: (1) the overabundance of monotypic genera, (2) an overabundance of species in relatively few genera, and (3) a significant, but undetermined number of the world species that are rare. The monotypes are where the chemical diversity lies. However, a large portion of the monotypes cannot be screened without a

high cost in acquisition, for perhaps 40% (5,000) of all genera of higher plants are rare. Initial collections will be mostly monotypes, but as collecting continues, an increasing percentage of species will belong to genera already collected, especially to those in Table 1.

Screening of genera might be limited to 7 species per genus (Table 5), but this point of decline in activity could shift to a higher number as new tumors are added to the prescreen, or to a lower number if activity in Table 5 is limited to certain tumors. The primary tumors were KB and PS that were employed for more than 20 yr and 12 yr, respectively. Perdue (1982) reported that the discovery of important antitumor agents had declined after 1967. The KB had been in use for approximately 7 yr and by the end of 1966, the USDA had accessioned about 12,000 samples; the first 10,000 samples represented 5,478 species in 2,075 genera (Perdue and Hartwell, 1969). Perhaps chemical diversity in plant samples had already reached a point of diminishing returns for KB screening. Genera with 7 or fewer species tested include 6,439 species; slightly more than half of all species are in genera with 11 or more species tested (Table 5).

The point of diminishing returns to random discovery of novel chemotherapeutic agents in higher plants depends also on phytogeographic sampling. Since the probability of discovering activity in a genus decreases after 7 species have been tested without success, I suggest 10,000 collections as a short-term goal to be represented by 500–1,000 species from each of Takhtajan's (1969) 12 floristic kingdoms or subkingdoms. Countries or states with a high diversity that correspond to the Takhtajan classification are: California, North Carolina, and Tien Shan in Kirgiz of Union of Soviet Socialist Republic (Boreal), Morocco (Tethyan), South Africa (South African), Zaire (African), Madagascar (Madagascan), Thailand and New Guinea (Indo-Malesian), Hawaii (Polynesian), New Caledonia (Neocaledonian), Mexico (Madrean), Brazil (Neotropical), Argentina (Antarctic) and finally Western Australia and Tasmania (Australian). As bioassays are evaluated and new ones evolve (Suffness and Douros, 1982), exploration should follow into new floristic regions until 85,000 species have been screened. The best chance for discovery of novel drugs from plants will be random screening with minimum duplication.

As a by-product of the NCI anticancer screening program, I propose to utilize listings of low priority taxa to develop a manual as a tool for identification of the world's most common plants. The potential utility for this is evident from Tables 1, 8, and Fig. 2. The 201 genera in Table 1 include ¼ of the world flora of higher plants (Table 3) and ⅓ of all species screened by the NCI. These common genera, combined with 2,905 common species (in 1,773 smaller genera), and 132 genera in Table 2, constitute 3,238 low priority taxa (SLOP) that occur frequently in ⅔ or more of the plants randomly collected within the stippled zone of Fig. 2 (e.g., Nicaragua and Malawi in Table 8). Genera in Table 2 were included with the common SLOP because their active agents have been discovered and were not of further interest to the NCI for KB and PS screening. However, it should be noted that compilations of common genera and species are incomplete due to termination of prescreening by the NCI. Minimally sampled (clear) regions in Fig. 2 probably have 20,000 species still easily available and acceptable to the NCI screen. By including areas where limited collecting occurred (striped), another 20,000 species may be possible.

Although a number of major tropical floras are nearing completion, or have recently been completed, their use in plant identification requires first the ability to classify an unknown plant to its family. Plant taxonomists learn family characteristics to facilitate identification, but the many exceptions to diagnostic characters of families make it difficult to comprehend their circumscriptions. On the other hand, generic concepts are more easily learned, and the ability to recognize common genera (Table 1) is clearly an advantage (Table 8, Fig. 2). Also, taxonomic keys to families and genera are primarily based on flowers and fruits, which are not always present at the time a botanist may need to collect and identify plants for pharmacological or ecological research. The development of this manual will put the emphasis on genera, especially with illustrated keys, descriptions noting key vegetative characters, comments on taxonomic status, and references.

Financial support for the development of the proposed manual is needed. A partnership, World Botanical Associates (WBA), was organized for this purpose. WBA is willing to collaborate with chemists, ecologists, and others to collect and/or identify plants for their research. By collecting plants for other purposes, such as for biological screening, much information is gained concerning a species' abundance and identifying characteristics. When I obtained re-collections for the NCI, a write-up or memorandum was prepared for each species that included a summary of field observations and literature on the plant's taxonomy, ecology, geography, economic, and folkloric uses. Farnsworth (1984) encourages botanists and chemists to collaborate, and from his experience, this kind of collaboration can improve the quality and productivity of research. This paper is an example, a product of collaboration between chemists and botanists.

With the growing concern to catalogue floras and preserve genetic diversity, especially where vegetation is rapidly being depleted (Diversity, 1982; Iltis, 1982; Science News, 1980), and the need to screen for novel drugs (NCI anticancer screening program; Farnsworth, 1984), create an identification manual for common plants (WBA objective), develop alternative crops and acquire wild germplasm of existing crops (high priority in ARS), perhaps plant exploration could be coordinated to serve the plurality of needs. Such coordination would seem especially valuable in view of the "lack of trained collectors" (USDA Report, 1981c) and stringent monetary policies that now exist.

ACKNOWLEDGMENTS

I am grateful to the National Cancer Institute (NCI) and to Dr. Matthew Suffness, Head of the Natural Products Branch of the NCI in Silver Spring, MD, for use of their microfilm reader and tapes, permission to publish this research, and for providing data on numbers of extracts tested of countries and states and other assistance. For reviews and comments, I thank Drs. Matthew Suffness, Jonathan Hartwell, retired since 1975 from Head of the Natural Products Section of the NCI, John Sawyer, Professor of Botany at Humboldt State University in Arcata, CA, and Daniel Norris, Professor of Botany at Humboldt State University. Additionally, I received many helpful suggestions from Charles Edson of World Botanical Associates, who was primarily inspirational in developing the phytogeography as a separate section of the paper, and from Dr. Herbert Gilliland, Jr., Assistant Professor of English at the U.S. Navel Academy in Annapolis, MD, who also provided many suggestions and much valued editorial assistance. Finally, I express my deepest appreciation to Judy A. Adams for her support and typing the manuscript.

## LITERATURE CITED


Agnew, A. D. Q. 1974. Upland Kenya Wild Flowers. A Flora of the Ferns and Herbaceous Flowering Plants of Upland Kenya, p. 573–610. Oxford Univ. Press, Ely House, London.

L'Association pour L'Etude Taxonomique de la Flore d'Afrique Tropicale. 1959. Vegetation map of Africa. With the assistance of UNESCO, Oxford Univ. Press, London.

Barclay, A. S., and R. E. Perdue, Jr. 1976. Distribution of anticancer activity in higher plants. Proc. 16th Annual Meeting, Soc. Econ. Bot., Plants and Cancer. Cancer Treatment Rep. 60: 1081–1113.

Beard, J. S. 1976. The Vegetation of the Dongara Area, Western Australia (data from p. 17). Vegmap Publications, Perth.

Bold, H. C. 1967. Morphology of Plants. 2nd ed. Harper & Row, New York.

Brenan, J. P. M. 1978. Some aspects of the phytogeography of tropical Africa. Ann. Missouri Bot. Gard. 65: 437–478.

Bruning, E. F. 1965. A guide and introduction to the vegetation of the Kerangas Forest and the Padangs of the Bako National Park, p. 289–318 (data from p. 291–300). Symposium on Ecological Research in Humid Tropics Vegetation, Government of Sarawak & UNESCO, Science Cooperation Office of Southeast Asia, Kuching.

Burbidge, N. T. 1960. The phytogeography of the Australian Region. Austral. J. Bot. 8: 75–211.

Cassady, J., and M. Suffness. 1980. Terpenoid antitumor agents. *In* J. Cassady and J. Douros, ed., Anticancer Agents Based on Natural Product Models, p. 201–269. Academic Press, New York.

Chapman, J. D., and F. White. 1970. The Evergreen Forests of Malawi (data from submontane and montane forests on p. 100–121). Commonw. Forest. Inst., Univ. Oxford.

Champion, H. G., and S. K. Seth. 1968. The Forest Types of India. A Revised Survey. Manager Govt. & Publications of India Press, Nasik & Delhi.

Cochrane, G. R. 1963. A physiognomic vegetation map of Australia. J. Ecol. 51: 639–655.

Croom, E. M., Jr. 1983. Documenting and evaluating herbal remedies. Econ. Bot. 37: 13–27.

Dale, I. R., and P. J. Greenway. 1961. Kenya Trees and Shrubs, p. 15–19. Buchanan's Kenya Estates, Nairobi.

Dasmann, R. F. 1973. A system for defining and classifying natural regions for purposes of conservation. International Union for Conservation of Nature and Natural Resources, Occasional Paper No. 7, Morges, Switzerland.

Davis, T. A. W., and P. W. Richards. 1934. The vegetation of Moraballi Creek, British Guiana: An ecological study of a limited area of tropical rain forest. Part II. J. Ecol. 22: 106–155.

Diversity, 1982. A news journal for the plant genetic resources community. D.G. Strauss, K. MacBain, D. Kalin, and P. Johannesen, ed. Diversity 1: 22 (Alarm over worldwide loss of genetic diversity voiced at State Department Conference).

Dodson, C. H., and A. H. Gentry. 1978. Flora of the Rio Palenque Science Center, Los Rios, Ecuador. Selbyana 4: 1–628.

Emberger, L. 1939. Apercu général sur la végétation du Maroc. *In* E. Rubel & W. Ludi, ed, Ergebnisse der Internationalen Pflanzengeographischen Exkursion durch Marokko und Westalgerien, 1936, p. 40–157 (data from p. 40–69). Verlag Hans Huber, Bern.

Farnsworth, N. R. 1984. How can the well be dry when it is filled with water? Econ. Bot. 38: 4–13.

Fraser, L., and J. W. Vickery. 1938. The ecology of the Upper Williams River and Barrington Tops Districts. II. Proc. Linn. Soc. New South Wales 63: 139–184 (data from Table 1).

Gentry, A. H. 1978. Floristic knowledge and needs in Pacific tropical America. Brittonia 30: 134–153.

Geran, R. I., N. H. Greenberg, M. M. Macdonald, A. M. Schumacher, and B. J. Abbott. 1972. Protocols for screening chemical agents and natural products against animal tumors and other biological systems, 3rd ed. Cancer Chemotherapy Rep., Part 3(3): 1–103.

Goldblatt, P. 1978. An analysis of the flora of southern Africa. Its characteristics, relationships, and origins. Ann. Missouri Bot. Gard. 65: 369–436.

Good, R. 1964. The Geography of the Flowering Plants. 3rd ed. Wiley, New York.

———. 1974. The Geography of the Flowering Plants. 4th ed. Longman, London.

Hartwell, J. L. 1976. Types of anticancer agents isolated from plants. Proc. 16th Annual Meeting, Soc. Econ. Bot., Plants and Cancer. Cancer Treatment Rep. 60: 1031–1067.

———. 1967–1971 (reprinted 1982 in book form by Quaterman, Lawrence, MA.). Plants used against cancer. A survey. Lloydia 30: 379–436, and 10 additional installments ending with 34: 386–437.

Hegnauer, R. 1977. The chemistry of the Compositae. *In* V. H. Heywood, J. B. Harborne, and B. L. Turner, ed, The Biology and Chemistry of the Compositae, Vol. 1, p. 283–335. Academic Press, London.

Holttum, R. E. 1973. Introduction to Pteridophyta. *In* J. C. Willis, A Dictionary of Flowering Plants and Ferns. 8th ed., revised by H. K. Airy Shaw, Cambridge Univ. Press.

Hueck, K. 1966. Die Walder Sudamerikas. *In* H. Walter, ed, Vegetationsmonographien der einzelnen GroBraume. Vol. II. G. Fischer Verlag, Stuttgart.

———, and P. Seibert. 1972. Vegetationskarte von Sudamerika. G. Fischer Verlag, Stuttgart.

Hutchinson, J. 1959. The Families of Flowering Plants. 2nd ed, Vol. 1, p. 242. Oxford Univ. Press, London.

Iltis, H. H. 1982. Discovery of No. 832: An essay in defense of the National Science Foundation. Desert Plants 2: 175–192.

Jackson, J. K. 1956. The vegetation of the Imatong Mountains, Sudan. J. Ecol. 44: 341–374 (data from Table 2, p. 353–355).

Jaffré, T., and M. Latham. 1974. Contribution a l'étude des relations sol-végétation sur un massif des roches ultrabasiques de la côte ouest de la Nouvelle-Caledonie: Le Boulinda. Adansonia, ser. 2, 14: 311–336 (data from p. 333–334).

Lemée, A. 1929–1943. Dictionnaire descriptif et synonymique des genres de plantes phanerogames. 8 vol. Imprimerie Commerciale et Administrative, Brest.

Letouzey, R. 1968. Étude Phytogéographique Du Cameroun (data from p. 167–173). Éditions Paul Lechevalier, Paris.

Man and Biosphere/Flora North America (MAB/FNA) Council. 1978. A Provisional Checklist of Species for Flora North America (revised). Monogr. Syst. Bot. 1, Flora North America Report 84, S. G. Shetler and L. E. Skog, ed. Missouri Bot. Gard., St. Louis, MO.

Mangenot, G. 1955. Étude sur les forêts des plaines et plateaux de la Côte-d'Ivoire. IV Études Éburneennes (data from tables on p. 23–29, 37, 45–46). Institut Francais D'Afrique Noire.

Mani, M. S. 1974. I. Introduction. II. Physical features. VI. The flora. *In* M. S. Mani, ed, Ecology and Biogeography in India, p. 1–10, 11–59, 159–177. W. Junk, The Hague.

Moldenke, H. N. 1973. Verbenaceae. *In* R. E. Woodson and R. W. Schery, Flora of Panama, Part IX. Ann. Missouri Bot. Gard. 60: 41–148.

Morello, J. 1958. La Provincia Fitogeographica del Monte. Opera Lilloana II, Tucuman.

Mullenders, W. 1954. La végétation de Kaniama (Entre-Lubishi-Lubilash, Congo Belge). Institut National pour l'Étude Agronomique du Congo Belge, Bruxelles, Ser. Sci. No. 61 (data from Table LVI).

Munz, P. A. 1959. A California Flora. In collaboration with D. D. Keck, p. 754–795. Univ. Calif. Press, Berkeley.

NCI (National Cancer Institute), HEW. 1974. Confirmed plant & animal materials, #62, Jan.

———. 1978. Plant Header File, Listed by family, genus (Feb. 14).

———, Memorandum. 1979a. M. Suffness to A. S. Barclay, Jan. 15. Genera extensively screened or completed (GESOC).

———, Memorandum. 1979b. M. Suffness to A. S. Barclay, June 19. Combined GESOC and 6-Plus list—Your memo of April 27, 1979.

———, Plant Header File. 1980. List of genera screened (Feb. 11).

———, Plant Header File. 1981. Materials confirmed active or collected 6 or more times. (April 8).

———, Plant Header Files. 1982. Microfilm tapes of "Plants arranged by genus; family" (June).

Nelson, C. J., J. N. Seiber, and L. P. Brower. 1981. Seasonal and intraplant variation of cardenolide content in the California milkweed, *Asclepias eriocarpa*, and implications for plant defense. J. Chem. Ecol. 7: 981–1010.

NIH (National Institutes of Health), HEW (Health, Education & Welfare), Memorandum. 1971. S. Richman to Distribution List, September 29. Minutes of Plant Meeting, September 22, 1971.

Perdue, R. E., Jr. 1976. Procurement of plant materials for antitumor screening. Proc. 16th Annual Meeting Soc. Econ. Bot., Plants and Cancer. Cancer Treatment Rep. 60: 987–998.

———. 1982. KB Cell Culture. I. Role in discovery of antitumor agents from higher plants. J. Nat. Prod. 45: 418–426.

———, and J. L. Hartwell. 1969. The search for plant sources of anticancer drugs. Morris Arbor. Bull. 20: 35–53.

Poore, M. E. D. 1968. Studies in Malaysian rain forests. I. The forest on Triassic sediments in Jengka Forest Reserve. J.Ecol. 56: 143–196 (data from Appendix, p. 189–196 and limited to species with 2 or more individuals recorded).

Prance, G. T. 1977. The phytogeographic subdivisions of Amazonia and their influence on the selection of biological reserves. *In* G. T. Prance and T. S. Elias, ed, Extinction Is Forever, p. 195–213. New York Botanical Garden, Bronx, NY.

Quézel, P. 1978. Analysis of the flora of Mediterranean and Saharan Africa. Ann. Missouri Bot. Gard. 65: 479–534.

Quisumbing, E. 1951. Medicinal Plants of the Philippines. Tech. Bull. 16, Philippines Dept. Agric. & Nat. Res., Manila.

Ramamoorthy, T. P. 1976. Acanthaceae. *In* C. J. Saldanha and D. H. Nicolson, ed, Flora of Hassan District, Karnataka, India, p. 538–561. Amerind Publ., New Delhi.

Raven, P. H. 1977. The California Flora. *In* M. G. Barbour and J. Major, ed, Terrestial Vegetation of California, p. 109–137. Wiley, New York.

Richards, P. W. 1963. Ecological notes on West African vegetation. III. The upland forest of Camerouns Mountain (data excludes "rare" species of herbs). J. Ecol. 51: 529–554.

Scagel, R. F., R. J. Bandoni, G. E. Rouse, W. B. Schofield, J. R. Stein, and T. M. C. Taylor. 1965. An Evolutionary Survey of the Plant Kingdom. Wadsworth, Belmont, CA.

Science News. 1980. Saving the Amazon, by J. Raloff and J. Silberner. 118: 218–221.

Shreve, F. 1964. Vegetation of the Sonoran Desert. *In* F. Shreve and I. L. Wiggins, Vegetation and Flora of the Sonoran Desert. Vol. 1, p. 26–126. Stanford Univ. Press, Palo Alto, CA.

———, and I. L. Wiggins. 1964. 2 vol. Vegetation and Flora of the Sonoran Desert. Stanford Univ. Press, Palo Alto, CA.

Spellman, D. L. 1975. Asclepiadaceae. *In* R. E. Woodson and R. W. Schery, Flora of Panama, Part VIII. Ann. Missouri Bot. Gard. 62: 103–156.

Spjut, R. W., and R. E. Perdue, Jr. 1976. Plant folklore: A tool for predicting sources of antitumor activity? Proc. 16th Annual Meeting, Soc. Econ. Bot., Plants and Cancer. Cancer Treatment Rep. 60: 979–985.

Statz, D., and F. B. Coon. 1976. Preparation of plant extracts for antitumor screening. Proc. 16th Annual Meeting, Soc. Econ. Bot., Plants and Cancer. Cancer Treatment Rep. 60: 999–1005.

Stevens, W. D. 1976. Asclepiadaceae. *In* C. J. Saldanha and D. H. Nicolson, ed, Flora of Hassan District, Karnataka, India, p. 437–455. Amerind Publ., New Delhi.

Suffness, M., and J. Douros. 1979. Drugs of plant origin. *In* V.T. DeVita and H. Busch, ed, Methods in Cancer Research, Vol. 16, p. 73–126. Academic Press, New York.

———, and ———. 1982. Current status of the NCI Plant and Animal Product Program. J. Nat. Prod. 45: 1–14.

Takhtajan, A. 1969. Flowering Plants. Origin and Dispersal. Translated by C. Jeffrey. Oliver & Boyd, Edinburgh.

Taylor, B. W. 1963. An outline of the vegetation of Nicaragua. J. Ecol. 51: 27–54 (data from p. 34–38).

Terrell, E. 1977. A Checklist of Names for 3,000 Vascular Plants of Economic Importance. Agric. Handbook No. 505, USDA, Washington, DC.

Udvardy, M. D. F. 1975. A classification of the biogeographical provinces of the world. International Union for Conservation of Nature and Natural Resources, Occasional Paper no. 18, Morges, Switzerland.

United States Department of Agriculture (USDA). 1970. Selected Weeds of the United States. Agric. Handbook No. 366, ARS, Washington, DC.

———. Plant Genetics & Germplasm Institute, Medicinal Plant Resources Laboratory, Memorandum. 1976. R. W. Spjut to A. S. Barclay, April 15. Patterns of anticancer activity in vascular plants.

———. Memorandum. 1978a. R. W. Spjut to A. S. Barclay, July 27. General plant samples collected in western U.S., 1978.

———. Memorandum. 1978b. R. W. Spjut to A. S. Barclay, Aug. 21. Plant genera to be sampled in the Sonoran Desert and western United States.

———. Memorandum. 1978c. R. W. Spjut to A. S. Barclay, Oct. 24. Plants Used Against Cancer: Sonoran Desert and Temperate North American genera with less than six collections tested.

———. Memorandum. 1979a. A. S. Barclay to M. Suffness, Jan. 25. Genera extensively screened or completed (GESOC).

———. Economic Botany Lab., Memorandum. 1979b. R. W. Spjut to A. S. Barclay, Oct. 5. Costs for recent field work to collect general samples and re-collections.

———. Memorandum. 1979c. R. W. Spjut to A. S. Barclay, Oct. 10. Quality comments on general samples from Nevada and California—September 1979.

———. Memorandum. 1980a. R. W. Spjut to POSI (Plants Of Special Interest File), Jan. 17. California annual to perennial genera not tested in the cancer program: Excludes grasses and woody plants.

———. 1980b. Combined list of genera extensively screened or completed (GESOC) and materials collected 6 or more times. Revised Feb. 14 (unpublished).

———. Travel Report, 1981a. Texas, Nevada, California and Oregon, R. W. Spjut, April 15–May 30.

———. Foreign Travel Report. 1981b. Western Australia and Tasmania, R. W. Spjut, Aug. 20–Nov. 8.

———. 1981c. The National Plant Germplasm System. I. Current Status (1980). II. Strengths. III. Long-Range Plan (1983–1997). Science & Education, Washington, DC.

Van Steenis, C. G. G. J. 1948–1949. Introduction, General considerations. *In* Flora Malesiana. Ser. 1, Vol. 4. Fig. 1. Delimitation and main divisions of the flora of Malaysia, on p. XI, & Fig. 45. Number of endemic genera of Phanerogams in the several islands and island groups of Malaysia, according to a census made in 1944, on p. LVI. Noordhoff, Leyden.

Watt, J. M., and M. Breyer-Brandwijk. 1962. The Medicinal and Poisonous Plants of Southern and Eastern Africa, 2nd ed. Livingstone, Edinburgh.

White, F. 1970. III. Floristics and plant geography. IV. Classification. *In* J. D. Chapman and F. White, The Evergreen Forests of Malawi, p. 38–112. Commonw. Forest. Inst., Oxford Univ. Press, London.

Willis, J. C. 1915. The endemic flora of Ceylon. Philos. Trans., Ser. B. 206: 307–342.

———. 1916. The distribution of species in New Zealand. Ann. Bot. 30: 437–457.

———. 1922. Age and Area. Cambridge Univ. Press, Cambridge.

———. 1966. A Dictionary of the Flowering Plants and Ferns. 7th ed, revised by H. K. Airy Shaw, Cambridge Univ. Press, Cambridge.

———. 1973. A Dictionary of the Flowering Plants and Ferns. 8th ed., revised by H. K. Airy Shaw, reprint 1980. Cambridge Univ. Press, Cambridge.

Zohary, M. 1973. Geobotanical Foundations of the Middle East. 2 vol. G. Fischer Verlag, Stuttgart.